

Multiple Access Gaussian Channels with Arbitrary Inputs: Optimal Precoding and Power Allocation

Samah A. M. Ghanem, *Member, IEEE*,

Abstract

In this paper, we derive new closed-form expressions for the gradient of the mutual information with respect to arbitrary parameters of the two-user multiple access channel (MAC). The derived relations generalize the fundamental relation between the derivative of the mutual information and the minimum mean squared error (MMSE) to multiuser setups. We prove that the derivative of the mutual information with respect to the signal to noise ratio (SNR) is equal to the MMSE plus a covariance induced due to the interference, quantified by a term with respect to the cross correlation of the multiuser input estimates, the channels and the precoding matrices. We also derive new relations for the gradient of the conditional and non-conditional mutual information with respect to the MMSE. Capitalizing on the new fundamental relations, we investigate the linear precoding and power allocation policies that maximize the mutual information for the two-user MAC Gaussian channels with arbitrary input distributions. We show that the optimal design of linear precoders may satisfy a fixed-point equation as a function of the channel and the input constellation under specific setups. We show also that the non-mutual interference in a multiuser setup introduces a term to the gradient of the mutual information which plays a fundamental role in the design of optimal transmission strategies, particularly the optimal precoding and power allocation, and explains the losses in the data rates. Therefore, we provide a novel interpretation of the interference with respect to the channel, power, and input estimates of the main user and the interferer.

Index Terms

Estimation Theory; Gradient of conditional mutual information; Gradient of non-conditional mutual information; Gradient of joint mutual information; Information Theory; Interference; MAC; MMSE; Multiuser I-MMSE; Mutual Information; Power Allocation; Precoding.

I. INTRODUCTION

LINEAR precoding for mutually interfering channels corrupted with independent Gaussian noise can be used to maximize the mutual information under a total power constraint [1]. In the multiple user setting, the mutual information is maximized by imposing a covariance structure on the Gaussian input vector that satisfies two conditions: First, transmitting along the right eigenvectors of the channel. Second, the power allocation follows the waterfilling based on the SNR in each parallel non-interacting sub-channel [2], [3]. The MAC channel stands as a special case of interference channels. With arbitrary inputs, the mutual information of the MAC channel is minimized due to the fact that at a certain point the inputs may lie in the null space of the channel, and this causes the mutual information to decay at such point. Therefore, the capacity-achieving strategies can be employed either at the transmitter side and/or the receiver side: by orthogonalizing the inputs using TDM/FDM techniques, or via using interference mitigation techniques, such as, dealing with interference as noise, rate splitting, successive interference cancellation, decoding interference, as well as receive diversity techniques [4], [5], [6], [7], [8], [9].

Precoding is another technique that can be used at the transmitter side to maximize the mutual information by exploiting the transmit diversity. On the one hand, precoding plays the role of power allocation diagonalizing the channel while, on the other hand, rotating the channel eigenvectors; thus shaping the constellation for better detection. Since practical considerations enforce the use of discrete constellations, such as PSK and QAM, which depart significantly from the optimal Gaussian distributions, it is of great importance to revisit the linear precoding problem under the constraint that each input follows a specified discrete constellation. In [3], [10], and [11], the optimal power allocation for parallel non-interacting channels with arbitrary inputs, and the optimal precoding for multiple inputs multiple outputs (MIMO) channels with different setups were obtained exploiting the relation between the mutual information and the MMSE [10], [12]. The results in [1] demonstrate that capacity-achieving strategies for Gaussian inputs may be suboptimal for discrete inputs. In [3] a general solution for the optimum precoder has been derived for the high and low-snr regimes capitalizing on first order expansions of the MMSE and relating

it to mutual information. The mercury/waterfilling solution for power allocation maximizes the mutual information compensating for the non-Gaussianity for the discrete constellations [13]. However, linear precoding techniques introduce rescaling among the channel inputs that achieves cleaner detection of the received signals and thus higher information rates. Therefore, this paper addresses the optimal precoding and optimal power allocation problem for multiuser channels driven by arbitrary inputs, extending previous works presented in [14], and [15], [16].

Due to the lack of explicit closed form expressions of the mutual information for binary inputs, which were provided only for the BPSK for the single input single output (SISO) case, [12], [17], [16], it is of particular importance to address connections between information theory and estimation theory to understand the communication framework under such inputs and to design optimal precoding and power allocation strategies that aim to maximize the reliable information rates.

In this paper¹, we first revisit the connections between the mutual information and the MMSE for the multiuser setup, to derive relations that account for the existence of an interferer channel. Therefore, the known fundamental relation between the derivative of the mutual information and the MMSE defined for point to point channels with any noise or input distributions in [12], is generalized to the multiuser case. Moreover, the relations for linear vector Gaussian channels in [10] is generalized for the multiuser channels where we extend these relations to the per-user gradient of the mutual information with respect to the MMSE, channels and precoders (power allocation) matrices of the user and the interferers. We then propose a linear precoder structure that aims to maximize the input output mutual information for the two-user MAC setting with channel state information known at both receiver and transmitter sides.

Throughout the paper, the following notation is employed, boldface uppercase letters denote matrices, lowercase letters denote scalars. The superscript, $(\cdot)^{-1}$, $(\cdot)^T$, $(\cdot)^*$, and $(\cdot)^\dagger$ denote inverse, transpose, conjugate and conjugate transpose operations. $(\cdot)^*$ denotes optimum. The (∇) denotes the gradient of a scalar function with respect to a variable. The $\mathbb{E}[\cdot]$ denotes the expectation operator. The $\|\cdot\|$ and $Tr\{\cdot\}$

¹Preliminary results of this work has been presented in WCSP [14].

denote the Euclidean norm, and the trace of a matrix, respectively.

The paper is organized as follows; section II introduces the system model and the problem formulation and, section III introduces the new fundamental relations between the mutual information and the MMSE. Section IV introduces the optimal precoding with arbitrary inputs and under different user setups. Section V introduces analysis at the asymptotic regime of low SNR. Section VI introduces the power allocation with arbitrary inputs and specializes it to the mercury/waterfilling under specific user setups. Section VII introduces simulation and analytical results where we introduce an interpretation of the interference, similar to known interpretations of the optimal power allocation over user main channels.

II. SYSTEM MODEL AND PROBLEM FORMULATION

Consider the deterministic complex-valued vector channel,

$$\mathbf{y} = \sqrt{snr} \mathbf{H}_1 \mathbf{P}_1 \mathbf{x}_1 + \sqrt{snr} \mathbf{H}_2 \mathbf{P}_2 \mathbf{x}_2 + \mathbf{n}, \quad (1)$$

where the $n_r \times 1$ dimensional vector \mathbf{y} and the $n_t \times 1$ dimensional vectors $\mathbf{x}_1, \mathbf{x}_2$ represent, respectively, the received vector and the independent zero-mean unit-variance transmitted information vectors from each user input to the MAC channel. The distributions of both inputs are not fixed, not necessarily Gaussian nor identical. The $n_r \times n_t$ complex-valued matrices $\mathbf{H}_1, \mathbf{H}_2$ correspond to the deterministic channel gains for both input channels (known to both encoder and decoder) and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ is the $n_r \times 1$ dimensional complex Gaussian noise with independent zero-mean unit-variance components. The optimization of the mutual information is carried out over all $n_t \times n_t$ precoding matrices $\mathbf{P}_1, \mathbf{P}_2$ that do not increase the transmitted power. The precoding problem can be cast as a constrained non-linear optimization problem as follows:

$$\max I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}) \quad (2)$$

Subject to:

$$\text{Tr} \left\{ \mathbb{E}_x [\mathbf{P}_1 \mathbf{x}_1 \mathbf{x}_1^\dagger \mathbf{P}_1^\dagger] \right\} = \text{Tr} \{ \mathbf{P}_1 \mathbf{P}_1^\dagger \} \leq Q_1 \quad (3)$$

$$\text{Tr} \left\{ \mathbb{E}_x [\mathbf{P}_2 \mathbf{x}_2 \mathbf{x}_2^\dagger \mathbf{P}_2^\dagger] \right\} = \text{Tr} \left\{ \mathbf{P}_2 \mathbf{P}_2^\dagger \right\} \leq Q_2 \quad (4)$$

To address the precoding and power allocation problems, we will revisit the fundamental relation between the mutual information and the MMSE with respect to the SNR for the multiuser case. The following section provides a set of new results that generalize of the known relations derived first by Guo, Shamai, and Verdu [12], and extends the results of linear vector channels by Palomar and Verdu [10] to the general case of multiuser channels.

III. NEW FUNDAMENTAL RELATIONS BETWEEN THE MUTUAL INFORMATION AND THE MMSE

The first contribution is given in the following theorem, which provides a generalization of the I-MMSE identity to the multiuser case and traverses back to the same identity of the single user case.

Theorem 1: The relation between the derivative of the mutual information with respect to the snr and the non-linear MMSE for a multiuser Gaussian channel satisfies:

$$\frac{dI(\text{snr})}{d\text{snr}} = \text{mmse}(\text{snr}) + \psi(\text{snr}) \quad (5)$$

Where,

$$\text{mmse}(\text{snr}) = \text{Tr} \left\{ \mathbf{H}_1 \mathbf{P}_1 \mathbf{E}_1 (\mathbf{H}_1 \mathbf{P}_1)^\dagger \right\} + \text{Tr} \left\{ \mathbf{H}_2 \mathbf{P}_2 \mathbf{E}_2 (\mathbf{H}_2 \mathbf{P}_2)^\dagger \right\}, \quad (6)$$

$$\begin{aligned} \psi(\text{snr}) = & \text{Tr} \left\{ \mathbf{H}_1 \mathbf{P}_1 \mathbb{E}_y [\mathbb{E}_{\mathbf{x}_1|y} [\mathbf{x}_1|y] \mathbb{E}_{\mathbf{x}_2|y} [\mathbf{x}_2|y]^\dagger] (\mathbf{H}_2 \mathbf{P}_2)^\dagger \right\} \\ & - \text{Tr} \left\{ \mathbf{H}_2 \mathbf{P}_2 \mathbb{E}_y [\mathbb{E}_{\mathbf{x}_2|y} [\mathbf{x}_2|y] \mathbb{E}_{\mathbf{x}_1|y} [\mathbf{x}_1|y]^\dagger] (\mathbf{H}_1 \mathbf{P}_1)^\dagger \right\}, \end{aligned}$$

Proof: The proof has been done for the two-user MAC channel as an example of a multiuser channel. See Appendix A. ■

The per-user MMSE is given respectively as follows:

$$\mathbf{E}_1 = \mathbb{E}_y [(\mathbf{x}_1 - \hat{\mathbf{x}}_1)(\mathbf{x}_1 - \hat{\mathbf{x}}_1)^\dagger] \quad (7)$$

$$\mathbf{E}_2 = \mathbb{E}_y [(\mathbf{x}_2 - \hat{\mathbf{x}}_2)(\mathbf{x}_2 - \hat{\mathbf{x}}_2)^\dagger]. \quad (8)$$

The input estimates of each user input is given respectively as follows:

$$\hat{\mathbf{x}}_1 = \mathbb{E}_{\mathbf{x}_1|y} [\mathbf{x}_1|y] = \sum_{\mathbf{x}_1, \mathbf{x}_2} \frac{\mathbf{x}_1 p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) p_{x_1}(\mathbf{x}_1) p_{x_2}(\mathbf{x}_2)}{p_y(\mathbf{y})} \quad (9)$$

$$\hat{\mathbf{x}}_2 = \mathbb{E}_{\mathbf{x}_2|y} [\mathbf{x}_2|y] = \sum_{\mathbf{x}_1, \mathbf{x}_2} \frac{\mathbf{x}_2 p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) p_{x_1}(\mathbf{x}_1) p_{x_2}(\mathbf{x}_2)}{p_y(\mathbf{y})}. \quad (10)$$

The conditional probability distribution of the Gaussian noise is defined as:

$$p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{\pi^{n_r}} e^{-\|\mathbf{y} - \sqrt{\text{snr}} \mathbf{H}_1 \mathbf{P}_1 \mathbf{x}_1 - \sqrt{\text{snr}} \mathbf{H}_2 \mathbf{P}_2 \mathbf{x}_2\|^2} \quad (11)$$

The probability density function for the received vector \mathbf{y} is defined as:

$$p_y(\mathbf{y}) = \sum_{\mathbf{x}_1, \mathbf{x}_2} p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) p_{x_1}(\mathbf{x}_1) p_{x_2}(\mathbf{x}_2). \quad (12)$$

Henceforth, the system MMSE with respect to the SNR is given by:

$$mmse(snr) = \mathbb{E}_{\mathbf{y}} \left[\left\| \mathbf{H}_1 \mathbf{P}_1 (\mathbf{x}_1 - \mathbb{E}_{\mathbf{x}_1|\mathbf{y}}[\mathbf{x}_1|\mathbf{y}]) \right\|^2 \right] + \mathbb{E}_{\mathbf{y}} \left[\left\| \mathbf{H}_2 \mathbf{P}_2 (\mathbf{x}_2 - \mathbb{E}_{\mathbf{x}_2|\mathbf{y}}[\mathbf{x}_2|\mathbf{y}]) \right\|^2 \right], \quad (13)$$

$$= Tr \left\{ \mathbf{H}_1 \mathbf{P}_1 \mathbf{E}_1 (\mathbf{H}_1 \mathbf{P}_1)^\dagger \right\} + Tr \left\{ \mathbf{H}_2 \mathbf{P}_2 \mathbf{E}_2 (\mathbf{H}_2 \mathbf{P}_2)^\dagger \right\} \quad (14)$$

as given in Theorem 1.

Note that the term $mmse(snr)$ is due to the users MMSEs, particularly, $mmse(snr) = mmse_1(snr) + mmse_2(snr)$ and $\psi(snr)$ are covariance terms that appear due to the covariance of the interferers. Those terms are with respect to the channels, precoders, and non-linear estimates of the user inputs. When the covariance terms vanish to zero, the mutual information with respect to the SNR will be equal to the mmse with respect to the SNR, this applies to the single user and point to point communications. Therefore, the result of Theorem 1 is a generalization of previous result and boils down to the result of Guo et al, [12] under certain conditions which are: (i) when the cross correlation between the inputs estimates equals zero (ii) when interference can be neglected (i.e. interference is very weak, very strong, or aligned) and (iii) under the single user setup.

Such generalized fundamental relation between the change in the mutliuser mutual information and the SNR is of particular relevance. Firstly, such result allows us to understand the behavior of per-user rates with respect to the interference due to the mutual interference and the interference due to other users behaviour in terms of power levels and channel strengths. In addition, the result allows us to be able to quantify the losses incurred due to the interference in terms of bits. Therefore, when the term $\psi(snr)$ equals zero. The derivative of the mutual information with respect to the SNR equals the total $mmse(snr)$:

$$\frac{dI(snr)}{dsnr} = mmse(snr), \quad (15)$$

which matches the result by Guo et. al in [12].

A. Remark on Interference Channels

The implication of the derived relation on the interference channel is of particular relevance. It is worth to note that we can capitalize on the new fundamental relation to extend the derivative with respect to the SNR to the conditional mutual information. To make this more clear, we capitalize on the chain rule of the mutual information which states the following:

$$I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}) = I(\mathbf{x}_1; \mathbf{y}) + I(\mathbf{x}_2; \mathbf{y}|\mathbf{x}_1) \quad (16)$$

Therefore, through this observation we can conclude the following theorem.

Theorem 2: The relation between the derivative of the conditional mutual information and the minimum mean squared error satisfies:

$$\frac{dI(\mathbf{x}_2; \mathbf{y}|\mathbf{x}_1)}{dsnr} = mmse_2(snr) + \psi(snr) \quad (17)$$

Proof: Taking the derivative of both sides of (16), and subtracting the derivative of $I(\mathbf{x}_1; \mathbf{y})$ which is equal to $mmse_1(\gamma snr)$, γ is a scaling factor, due to the fact that x_1 is decoded first considering the other users' input x_2 as noise. Therefore, Theorem 2 has been proved. ■

Note that the derivative of the conditional mutual information as well as the non-conditional mutual information can be scaled due to different SNRs in a two-user interference channel. However, the scaling is straightforward to apply. For further details, refer to [16], [18], and [19].

To solve the optimization problem in (2) subject to (3) and (4), we need to know the relation between the mutual information and the non-linear MMSE in the context of MAC channels. The following theorem puts forth a new fundamental relation between the mutual information and the non-linear MMSE for MIMO MAC channels ². In particular, the gradient of the mutual information with respect to arbitrary parameters has been derived. First we will present the gradient of the mutual information, with respect to the channel in Theorem 3, then with respect to the precoder (or power allocation) in Theorem 4.

²Noice that through the flow of the paper, we use MAC channel instead of MIMO MAC since such relations are general and apply even when users are transmitting over scalar channels.

The following theorems in addition to Theorem 1 generalizes the connections between information theory and estimation theory to the multiuser case.

Theorem 3: The relation between the gradient of the mutual information with respect to the channel and the non-linear MMSE for a two user-MAC channel with arbitrary inputs (1) satisfies:

$$\nabla_{\mathbf{H}_1} I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}) = \mathbf{H}_1 \mathbf{P}_1 \mathbf{E}_1 \mathbf{P}_1^\dagger - \mathbf{H}_2 \mathbf{P}_2 \mathbb{E}[\hat{\mathbf{x}}_2 \hat{\mathbf{x}}_1^\dagger] \mathbf{P}_1^\dagger \quad (18)$$

$$\nabla_{\mathbf{H}_2} I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}) = \mathbf{H}_2 \mathbf{P}_2 \mathbf{E}_2 \mathbf{P}_2^\dagger - \mathbf{H}_1 \mathbf{P}_1 \mathbb{E}[\hat{\mathbf{x}}_1 \hat{\mathbf{x}}_2^\dagger] \mathbf{P}_2^\dagger \quad (19)$$

Proof: See Appendix B. ■

Theorem 4: The relation between the gradient of the mutual information with respect to the precoding matrix and the non-linear MMSE for a two user-MAC channel with arbitrary inputs (1) satisfies:

$$\nabla_{\mathbf{P}_1} I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}) = \mathbf{H}_1^\dagger \mathbf{H}_1 \mathbf{P}_1 \mathbf{E}_1 - \mathbf{H}_1^\dagger \mathbf{H}_2 \mathbf{P}_2 \mathbb{E}[\hat{\mathbf{x}}_2 \hat{\mathbf{x}}_1^\dagger] \quad (20)$$

$$\nabla_{\mathbf{P}_2} I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}) = \mathbf{H}_2^\dagger \mathbf{H}_2 \mathbf{P}_2 \mathbf{E}_2 - \mathbf{H}_2^\dagger \mathbf{H}_1 \mathbf{P}_1 \mathbb{E}[\hat{\mathbf{x}}_1 \hat{\mathbf{x}}_2^\dagger] \quad (21)$$

Proof: See Appendix C. ■

Theorems 3 and 4 provides intuitions about the change of the mutual information with respect to the changes in the channel or the precoding (power allocation). There is a straightforward connection between both changes when each gradient is scaled with respect to the changing arbitrary parameter, we can write this connection as follows:

$$\nabla_{\mathbf{P}_1} I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}) \mathbf{P}_1^\dagger = \mathbf{H}_1^\dagger \nabla_{\mathbf{H}_1} I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}), \quad (22)$$

$$\nabla_{\mathbf{P}_2} I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}) \mathbf{P}_2^\dagger = \mathbf{H}_2^\dagger \nabla_{\mathbf{H}_2} I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}), \quad (23)$$

Note that we can derive the gradient of the mutual information with respect to any arbitrary parameter following similar steps of the proof of the previous two theorems. Note also that the derived relations in (18), (19), (20), and (21) reduce to the relation between the gradient of the mutual information and the non-linear MMSE derived for the linear vector Gaussian channels [10] if the cross correlation between the input estimates is zero. This is possible when the users inputs are orthogonal or interference alignment

is performed. In addition, the derived relation between the gradient of the mutual information and the MMSE of the two-user MAC setting in Theorem 3 will lead to a new formulation of the input non-linear estimates, that is discussed in the following Theorem.

Theorem 5: The estimates of the inputs \mathbf{x}_1 and \mathbf{x}_2 of the two-user MAC channel with arbitrary inputs given the output \mathbf{y} can be expressed as:

$$\mathbf{H}_1 \mathbf{P}_1 \mathbb{E}_{x_1|y}[\mathbf{x}_1|\mathbf{y}] + \mathbf{H}_2 \mathbf{P}_2 \mathbb{E}_{x_2|y}[\mathbf{x}_2|\mathbf{y}] = \mathbf{y} + \frac{\nabla_y p_y(\mathbf{y})}{p_y(\mathbf{y})} \quad (24)$$

Proof: See Appendix D. ■

(24) can be solved numerically to obtain the estimates of the inputs. However, we can extract a special case of Theorem 5 in the following corollary.

Corollary 1: If the channel is a linear vector Gaussian channel, Theorem 5 leads to the following formulation:

$$\mathbb{E}_{x|y}[\mathbf{x}|\mathbf{y}] = \left(\mathbf{y} + \frac{\nabla_y p_y(\mathbf{y})}{p_y(\mathbf{y})} \right) (\mathbf{H}\mathbf{P})^{-1} \quad (25)$$

(25) can be manipulated in terms of its zero forcing linear counter measure, where the distance between a linear and a non-linear estimate is approximately equal to ϵ , where $\epsilon = (\frac{\nabla_y p_y(\mathbf{y})}{p_y(\mathbf{y})})(\mathbf{H}\mathbf{P})^{-1}$. Further, the non-linear estimates given in (9) and (10) give a statistical intuition to the problem. However, the most used in practical setups are linear estimation sub-optimal methods, where linear estimates can be found via the linear MMSE. In particular, the input estimates can be found by deriving the optimal Wiener receive filters solving a minimization optimization problem of the mean squared error (MSE). The following theorem provides the linear MMSE estimates.

Theorem 6: The linear estimates of the inputs \mathbf{x}_1 , \mathbf{x}_2 of the two-user MAC channel given the output \mathbf{y} can be expressed as:

$$\hat{\mathbf{x}}_1 = \mathbf{P}_1^\dagger \mathbf{H}_1^\dagger (\mathbf{I} + \mathbf{P}_1^\dagger \mathbf{H}_1^\dagger \mathbf{P}_1 \mathbf{H}_1 + \mathbf{P}_2^\dagger \mathbf{H}_2^\dagger \mathbf{P}_2 \mathbf{H}_2)^{-1} \mathbf{y} \quad (26)$$

$$\hat{\mathbf{x}}_2 = \mathbf{P}_2^\dagger \mathbf{H}_2^\dagger (\mathbf{I} + \mathbf{P}_1^\dagger \mathbf{H}_1^\dagger \mathbf{P}_1 \mathbf{H}_1 + \mathbf{P}_2^\dagger \mathbf{H}_2^\dagger \mathbf{P}_2 \mathbf{H}_2)^{-1} \mathbf{y} \quad (27)$$

Proof: See Appendix E. ■

Worth to note that the linear estimation process of the inputs corresponds to multiplying the user Wiener MMSE receive filters to the received vector. Note that each receive filter contains the precoding (or power allocation), the user channel, in addition to a structure of the linear MMSE, that is, the inverse of the covariance structure of the received vector.

B. Gradient of the Conditional and Non-Conditional Mutual Information

Theorem 4 shows how much rate is lost due to the other user, this is due to the fact that some terms in the gradient of the mutual information preclude the effect of the mutual interference of the main links. Therefore, we can account for such quantified rate loss via optimal power allocation and optimal precoding. Then, in order to be able to understand the achieved rates of each user in a MAC channel, we capitalize on the chain rule of the mutual information to derive the conditional mutual information as follows,

$$I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}) = I(\mathbf{x}_2; \mathbf{y}) + I(\mathbf{x}_1; \mathbf{y}|\mathbf{x}_2) \quad (28)$$

and,

$$I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}) = I(\mathbf{x}_1; \mathbf{y}) + I(\mathbf{x}_2; \mathbf{y}|\mathbf{x}_1) \quad (29)$$

Clearly, we know that $I(\mathbf{x}_2; \mathbf{y})$ is the mutual information when user 2 is decoded first considering user 1 signal as noise. Therefore, we can write it as follows,

$$I(\mathbf{x}_2; \mathbf{y}) = \left[\log \frac{p_{y|x_2}(\mathbf{y}|\mathbf{x}_2)}{p_y(\mathbf{y})} \right] \quad (30)$$

$$p_{y|x_2}(\mathbf{y}|\mathbf{x}_2) = \frac{1}{\pi^{n_r}} e^{-(\mathbf{y} - \sqrt{\text{snr}} \mathbf{H}_2 \mathbf{P}_2 \mathbf{x}_2)^\dagger (\mathbf{P}_1^\dagger \mathbf{H}_1^\dagger \mathbf{H}_1 \mathbf{P}_1 + \mathbf{I})^{-1} (\mathbf{y} - \sqrt{\text{snr}} \mathbf{H}_2 \mathbf{P}_2 \mathbf{x}_2)} \quad (31)$$

$$p_y(\mathbf{y}) = \sum_{x_2} p_{y|x_2}(\mathbf{y}|\mathbf{x}_2) p_{x_2}(\mathbf{x}_2) \quad (32)$$

Based on such definition, we conclude the following theorem which provides a new fundamental relation between the gradient of the mutual information with respect to the precoder of the other user given that the other user will be secondly decoded.

Theorem 7: The gradient of the mutual information with respect to the precoder, for a scaled user power when the other user input is considered as noise is as follows,

$$\nabla_{\mathbf{P}_1} I(\mathbf{x}_2; \mathbf{y}) = \mathbf{H}_2 \mathbf{P}_2 \mathbf{E}_2 \mathbf{P}_2^\dagger \mathbf{H}_2^\dagger \mathbf{H}_1^\dagger \mathbf{H}_1 \mathbf{P}_1 (\mathbf{P}_1^\dagger \mathbf{H}_1^\dagger \mathbf{H}_1 \mathbf{P}_1 + \mathbf{I})^{-1} \quad (33)$$

$$\nabla_{\mathbf{P}_2} I(\mathbf{x}_1; \mathbf{y}) = \mathbf{H}_1 \mathbf{P}_1 \mathbf{E}_1 \mathbf{P}_1^\dagger \mathbf{H}_1^\dagger \mathbf{H}_2^\dagger \mathbf{H}_2 \mathbf{P}_2 (\mathbf{P}_2^\dagger \mathbf{H}_2^\dagger \mathbf{H}_2 \mathbf{P}_2 + \mathbf{I})^{-1} \quad (34)$$

Proof: The proof follows similar steps of the proof of Theorem 4. ■

When each user is transmitting over a single channel, the new relation in Theorem 7 will be more clearly understood in terms of the effect of the interference plus noise power scaling on the gradient, see [18].

Corollary 2: The gradient of the conditional mutual information will be as follows,

$$\nabla_{\mathbf{P}_1} I(\mathbf{x}_1; \mathbf{y} | \mathbf{x}_2) = \nabla_{\mathbf{P}_1} I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}) - \nabla_{\mathbf{P}_1} I(\mathbf{x}_2; \mathbf{y}) \quad (35)$$

$$\nabla_{\mathbf{P}_2} I(\mathbf{x}_2; \mathbf{y} | \mathbf{x}_1) = \nabla_{\mathbf{P}_2} I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}) - \nabla_{\mathbf{P}_2} I(\mathbf{x}_1; \mathbf{y}) \quad (36)$$

Proof: The proof of the corollary follows from the chain rule of mutual information and the gradient of the mutual information derived for the sum rate and non-conditional rates. ■

IV. OPTIMAL PRECODING WITH ARBITRARY INPUTS

Capitalizing on the relation between the gradient of the mutual information with respect to the precoding matrix and the MMSE, we can solve the optimization problem in (2) subject to (3) and (4). The following theorem provides the optimal precoding of the two-user MAC channel with arbitrary inputs which can be also specialized to Gaussian inputs, see [16], [18], [15].

Theorem 8: Let the distribution of the arbitrary inputs $\mathbf{x}_1, \mathbf{x}_2$ to the channel in (1) be $\mathbf{p}_{\mathbf{x}_1}(\mathbf{x}_1), \mathbf{p}_{\mathbf{x}_2}(\mathbf{x}_2)$, respectively. The optimal precoding matrices satisfy the following:

$$\mathbf{P}_1^* = \nu^{-1} \mathbf{H}_1^\dagger \mathbf{H}_1 \mathbf{P}_1^* \mathbf{E}_1 - \nu^{-1} \mathbf{H}_1^\dagger \mathbf{H}_2 \mathbf{P}_2^* \mathbb{E}[\hat{\mathbf{x}}_2 \hat{\mathbf{x}}_1^\dagger] \quad (37)$$

$$\mathbf{P}_2^* = \nu^{-1} \mathbf{H}_2^\dagger \mathbf{H}_2 \mathbf{P}_2^* \mathbf{E}_2 - \nu^{-1} \mathbf{H}_2^\dagger \mathbf{H}_1 \mathbf{P}_1^* \mathbb{E}[\hat{\mathbf{x}}_1 \hat{\mathbf{x}}_2^\dagger] \quad (38)$$

Where ν are the received *snr* normalized by the Lagrange multipliers of the optimization problem in (2).

Proof: The proof of Theorem 8 relies on the KKT conditions [20] and the relation between the gradient of the mutual information with respect to the precoding matrices and the MMSE, see Appendix F. ■

There are unique \mathbf{P}_1^* , \mathbf{P}_2^* that satisfy the KKT conditions when the problem is strictly concave, corresponding to the global maximum. Using Monte-Carlo method, we can compute the MMSE matrices (7) and (8) for discrete constellations. To obtain the solution of (37) and (38), we can use an iterative approach, similar to [1], [3], and [11]. It is very important to notice that the optimal precoders of (37) and (38) satisfies a fixed point equation under two specific setups:

First, for the multiple-channels-per user in the two-user-MAC if and only if the second term in the gradient of the mutual information with respect to the precoding matrix is zero, this occurs when both inputs are orthogonal, or in this case, we can think about it as if the second user multiple-inputs have no influence to the first user multiple inputs, therefore, each user precoding is over his mutually interfering channels only. If each user is using OFDM signaling, i.e., transmitting over parallel independent Gaussian channels, the problem breaks into two separate problems, and the precoding solution for each is a fixed point equation.

Second, for the two-user-MAC case with each user is mutually interfering with the other, in this case the relation between the gradient of the mutual information with respect to the precoding matrix can be manipulated as in [10], thus, the precoder structure is similar to the one introduced in [3], [15]. We define in the following theorem a general form of the optimal precoder's structure for the two-user MAC channel in (37) and (38) under the two specific setups where the precoder would satisfy a fixed point equation, therefore, the solution applies to the low-snr and high-snr regimes. However, it also reduces to the optimal power allocation solution when the precoding is precluded as will be shown later. Therefore, the precoder should admit a structure that performs matching of the strongest source modes to the weakest

noise modes, and this alignment enforces permutation process to appear in the diagonal power allocation setup.

Theorem 9: The non-unique first-order optimal precoders that maximize the mutual information for a two-user MAC subject to per-user power constraint, and under the assumption that the second term in (37), and (38) equals zero, can be given as follows:

$$\mathbf{P}_1^* = \mathbf{U}_1 \mathbf{D}_1 \mathbf{R}_1^\dagger \quad (39)$$

$$\mathbf{P}_2^* = \mathbf{U}_2 \mathbf{D}_2 \mathbf{R}_2^\dagger \quad (40)$$

With $\mathbf{U}_1, \mathbf{U}_2$ are unitary matrices, $\mathbf{D}_1, \mathbf{D}_2$ are diagonal matrices, and $\mathbf{R}_1, \mathbf{R}_2$ are rotation matrices.

Proof: The theorem follows the relation between the gradient of the mutual information with respect to the precoding matrix and the decomposition of the components of the channel, precoder, and MMSE matrices, and is a direct consequence to [Theorem 1, 5], see Appendix G. ■

The optimal precoders are as follows, $\mathbf{P}_1 = \mathbf{U}_1 \mathbf{D}_1 \mathbf{R}_1^\dagger$ and $\mathbf{P}_2 = \mathbf{U}_2 \mathbf{D}_2 \mathbf{R}_2^\dagger$ with $\mathbf{U}_1 = \mathbf{V}_{\mathbf{H}_1}, \mathbf{U}_2 = \mathbf{V}_{\mathbf{H}_2}$, correspond to the per user channel right singular vectors, respectively.

$\mathbf{D}_1 = \text{diag}(\sqrt{\mathbf{p}_{1,1}}, \dots, \sqrt{\mathbf{p}_{1,n_t}})$ and $\mathbf{D}_2 = \text{diag}(\sqrt{\mathbf{p}_{2,1}}, \dots, \sqrt{\mathbf{p}_{2,n_t}})$ are power allocation matrices, such that $\mathbf{p}_{l,m}$ is admitted by user l over sub/channel m ; this in fact corresponds to the mercury/waterfilling. $\mathbf{R}_1 = \mathbf{\Pi} \mathbf{U}_{\mathbf{E}_1}$ and $\mathbf{R}_2 = \mathbf{\Pi} \mathbf{U}_{\mathbf{E}_2}$ are rotation matrices that contain in their structure the eigenvectors of the user MMSE matrix which can be permuted and/or projected with $\mathbf{\Pi}$ based on the correlation of the inputs and their estimates. Therefore, the rotation matrix insures firstly, allocation of power into the strongest channel singular vectors, and secondly, diagonalizes the MMSE matrix to insure un-correlating the error or in other words independence between inputs, enforcing the setup defined in (37) and (38) for the two special setups discussed previously about each user inputs, i.e., when the interference from each user to the other is cancelled or precluded, and approximated by zero.

V. THE LOW-SNR REGIME

We now consider the optimal precoding policy for the two-user MAC Gaussian channel with arbitrary input distributions in the regime of low-snr. Consider a zero-mean uncorrelated complex inputs, with $\mathbb{E}[\mathbf{x}_1 \mathbf{x}_1^\dagger] = \mathbb{E}[\mathbf{x}_2 \mathbf{x}_2^\dagger] = \mathbf{I}$, and $\mathbb{E}[\mathbf{x}_1 \mathbf{x}_1^T] = \mathbb{E}[\mathbf{x}_2 \mathbf{x}_2^T] = \mathbf{0}$. We consider the low-snr expansion to the MMSE of equation (13). Note that it can be easily deduced that the Taylor expansion of the non-linear MMSE in (13) will lead to the first order Taylor expansion of the linear MMSE for the Gaussian inputs setup. Thus, the low-snr expansion of the MMSE matrix can be expressed as:

$$\mathbf{E} = \mathbf{I} - (\mathbf{H}_1 \mathbf{P}_1)^\dagger \mathbf{H}_1 \mathbf{P}_1 .snr - (\mathbf{H}_2 \mathbf{P}_2)^\dagger \mathbf{H}_2 \mathbf{P}_2 .snr + \mathcal{O}(snr^2), \quad (41)$$

with $\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2$. Consequently,

$$mmse(snr) = Tr \left\{ \mathbf{H}_1 \mathbf{P}_1 \mathbf{E}_1 (\mathbf{H}_1 \mathbf{P}_1)^\dagger \right\} + Tr \left\{ \mathbf{H}_2 \mathbf{P}_2 \mathbf{E}_2 (\mathbf{H}_2 \mathbf{P}_2)^\dagger \right\} \quad (42)$$

$$\begin{aligned} &= Tr \left\{ \mathbf{H}_1 \mathbf{P}_1 (\mathbf{H}_1 \mathbf{P}_1)^\dagger \right\} + Tr \left\{ \mathbf{H}_2 \mathbf{P}_2 (\mathbf{H}_2 \mathbf{P}_2)^\dagger \right\} \\ &\quad - Tr \left\{ (\mathbf{H}_1 \mathbf{P}_1 (\mathbf{H}_1 \mathbf{P}_1)^\dagger)^2 \right\} .snr - Tr \left\{ (\mathbf{H}_2 \mathbf{P}_2 (\mathbf{H}_2 \mathbf{P}_2)^\dagger)^2 \right\} .snr + \mathcal{O}(snr^2). \end{aligned} \quad (43)$$

Note that due to our new result of Theorem 1, we cannot apply immediately the fundamental relationship between mutual information and MMSE in [10], [12]. Therefore, applying our new result, the low-snr Taylor expansion of the mutual information is given by:

$$\begin{aligned} I(snr) &= Tr \left\{ \mathbf{H}_1 \mathbf{P}_1 (\mathbf{H}_1 \mathbf{P}_1)^\dagger \right\} .snr + Tr \left\{ \mathbf{H}_2 \mathbf{P}_2 (\mathbf{H}_2 \mathbf{P}_2)^\dagger \right\} .snr \\ &\quad - Tr \left\{ (\mathbf{H}_1 \mathbf{P}_1 (\mathbf{H}_1 \mathbf{P}_1)^\dagger)^2 \right\} .snr^2 - Tr \left\{ (\mathbf{H}_2 \mathbf{P}_2 (\mathbf{H}_2 \mathbf{P}_2)^\dagger)^2 \right\} .snr^2 + \\ &\quad + Tr \left\{ \mathbf{H}_1 \mathbf{P}_1 (\mathbf{H}_1 \mathbf{P}_1)^\dagger \mathbf{H}_2 \mathbf{P}_2 (\mathbf{H}_2 \mathbf{P}_2)^\dagger \right\} .snr^2 \\ &\quad - Tr \left\{ \mathbf{H}_2 \mathbf{P}_2 (\mathbf{H}_2 \mathbf{P}_2)^\dagger \mathbf{H}_1 \mathbf{P}_1 (\mathbf{H}_1 \mathbf{P}_1)^\dagger \right\} .snr^2 + \mathcal{O}(snr^3) \end{aligned} \quad (44)$$

The wideband slope – which indicates how fast the capacity is achieved in terms of required bandwidth – is inversely proportional to the second order terms of the mutual information in the low-snr Taylor expansion (44). Therefore, this term is a key low-power performance measure since the bandwidth required

to sustain a given rate with a given low power, i.e., minimal energy per bit, is inversely proportional to this term [21].

According to (44), for first-order optimality, subject to (3) and (4), the form of the optimal precoder follows from the low-snr expansions the form of optimal precoder for the complex Gaussian inputs settings.

To prove this claim, lets re-define our optimization problem as follows:

$$\max \quad Tr \left\{ \mathbf{H}_1 \mathbf{P}_1 (\mathbf{H}_1 \mathbf{P}_1)^\dagger \right\} .snr + Tr \left\{ \mathbf{H}_2 \mathbf{P}_2 (\mathbf{H}_2 \mathbf{P}_2)^\dagger \right\} .snr \quad (45)$$

Subject to:

$$Tr \left\{ \mathbf{P}_1 \mathbf{P}_1^\dagger \right\} \leq 1 \quad (46)$$

$$Tr \left\{ \mathbf{P}_2 \mathbf{P}_2^\dagger \right\} \leq 1 \quad (47)$$

Lets do an eigen value decomposition such that:

$$\mathbf{H}_1 \mathbf{H}_1^\dagger = \mathbf{U}_{\mathbf{H}_1 \mathbf{H}_1^\dagger} \mathbf{\Lambda}_{\mathbf{H}_1 \mathbf{H}_1^\dagger} \mathbf{U}_{\mathbf{H}_1 \mathbf{H}_1^\dagger}^\dagger, \quad (48)$$

$$\mathbf{H}_2 \mathbf{H}_2^\dagger = \mathbf{U}_{\mathbf{H}_2 \mathbf{H}_2^\dagger} \mathbf{\Lambda}_{\mathbf{H}_2 \mathbf{H}_2^\dagger} \mathbf{U}_{\mathbf{H}_2 \mathbf{H}_2^\dagger}^\dagger \quad (49)$$

Let:

$$\tilde{\mathbf{P}}_1 = \mathbf{U}_{\mathbf{H}_1 \mathbf{H}_1^\dagger}^\dagger \mathbf{P}_1 \quad (50)$$

$$\tilde{\mathbf{P}}_2 = \mathbf{U}_{\mathbf{H}_2 \mathbf{H}_2^\dagger}^\dagger \mathbf{P}_2 \quad (51)$$

Let $\mathbf{Z}_1 \succeq \mathbf{0}$ and $\mathbf{Z}_2 \succeq \mathbf{0}$ are positive semi-definite, such that:

$$\mathbf{Z}_1 = \tilde{\mathbf{P}}_1 \tilde{\mathbf{P}}_1^\dagger \quad (52)$$

$$\mathbf{Z}_2 = \tilde{\mathbf{P}}_2 \tilde{\mathbf{P}}_2^\dagger \quad (53)$$

Substitute (48) to (53) into (45), (46), and (47), the optimization problem can be re-written as:

$$\max \quad Tr \left\{ \mathbf{Z}_1 \mathbf{\Lambda}_{\mathbf{H}_1 \mathbf{H}_1^\dagger} \right\} .snr + Tr \left\{ \mathbf{Z}_2 \mathbf{\Lambda}_{\mathbf{H}_2 \mathbf{H}_2^\dagger} \right\} .snr \quad (54)$$

Subject to:

$$Tr \left\{ \mathbf{Z}_1 \right\} \leq 1 \quad (55)$$

$$\text{Tr}\{\mathbf{Z}_2\} \leq 1 \quad (56)$$

Solving the KKT conditions for the Lagrangian of the objective function (54) subject to (55) and (56) leads to:

$$\mathbf{Z}_1 = \lambda^{-1} \mathbf{\Lambda}_{\mathbf{H}_1 \mathbf{H}_1^\dagger} \cdot \text{snr}, \quad (57)$$

$$\mathbf{Z}_2 = \lambda^{-1} \mathbf{\Lambda}_{\mathbf{H}_2 \mathbf{H}_2^\dagger} \cdot \text{snr} \quad (58)$$

Where λ is the Lagrange multiplier of the optimization problem. The results in (57) and (58) prove that the optimal precoders in the low-snr perform mainly two operations: Firstly, it aligns the transmit directions with the eigenvectors of each sub-channel. Secondly, it performs power allocation over the sub-channels.

VI. OPTIMAL POWER ALLOCATION WITH ARBITRARY INPUTS

To derive the optimal power allocation for MAC Gaussian channels driven by arbitrary inputs, we need to redefine the same objective function subject to the per user power constraints. The power allocation is a special case of the precoding which inherently includes a power allocation operation. However, the first reshapes the constellation by rotating or redefining the minimum distance between the points, while the second redistributes the power over the channels to guarantee a better utilization of the channel under the defined design criterion. Therefore, the power allocation problem can be cast as a constrained non-linear optimization problem as follows:

$$\max I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}) \quad (59)$$

Subject to:

$$(\mathbf{P}_1)_{ij} = \sqrt{p_{1j}}, \quad (\mathbf{P}_2)_{ij} = \sqrt{p_{2j}} \quad (60)$$

$$\sum_{j=1}^{n_t} p_{1j} \leq 1, \quad p_{1j} \geq 0 \quad \forall j = 1, \dots, n_t \quad (61)$$

$$\sum_{j=1}^{n_t} p_{2j} \leq 1, \quad p_{2j} \geq 0 \quad \forall j = 1, \dots, n_t \quad (62)$$

where $\mathbf{p}_1 = \text{diag}(\sqrt{p_{1,1}}, \dots, \sqrt{p_{1,n_t}})$, and $\mathbf{p}_2 = \text{diag}(\sqrt{p_{2,1}}, \dots, \sqrt{p_{2,n_t}})$. Capitalizing on the relation between the gradient of the mutual information with respect to the power allocation matrix and the MMSE, we provide the optimal power allocation in the following theorem.

Theorem 10: The optimal power allocation that solves (59) subject to (60) to (62) satisfies:

$$\mathbf{p}_1^* = \gamma_1^{-1} \left(\mathbf{P}_1^* \mathbf{H}_1^\dagger \mathbf{H}_1 \mathbf{P}_1^* \mathbf{E}_1 - \mathbf{P}_1^* \mathbf{H}_1^\dagger \mathbf{H}_2 \mathbf{P}_2^* \mathbb{E}_{\mathbf{y}}[\hat{\mathbf{x}}_2 \hat{\mathbf{x}}_1^\dagger] \right)_{1j} \quad (63)$$

$$\mathbf{p}_2^* = \gamma_2^{-1} \left(\mathbf{P}_2^* \mathbf{H}_2^\dagger \mathbf{H}_2 \mathbf{P}_2^* \mathbf{E}_2 - \mathbf{P}_2^* \mathbf{H}_2^\dagger \mathbf{H}_1 \mathbf{P}_1^* \mathbb{E}_{\mathbf{y}}[\hat{\mathbf{x}}_1 \hat{\mathbf{x}}_2^\dagger] \right)_{2j} \quad (64)$$

Where γ_1 and γ_2 are the Lagrange multipliers normalized by the received *snr*, of the optimization problem in (59).

Proof: The proof of Theorem 10 relies on the KKT conditions [20] and the relation between the gradient of the mutual information with respect to the precoding matrices and the MMSE, see Appendix H. ■

For the special case where we have a single channel per user in the two-user MAC, (63) and (64) can be simplified to the following:

$$\mathbf{p}_1^* = \gamma_1^{-1} p_{1j}^* |h_{1j}|^2 \mathbb{E}_{\mathbf{y}}[(\mathbf{x}_1 - \hat{\mathbf{x}}_1)(\mathbf{x}_1 - \hat{\mathbf{x}}_1)^\dagger] - \gamma_1^{-1} p_{2j}^* |h_{1j}|^2 |h_{2j}| \mathbb{E}_{\mathbf{y}}[(\mathbf{x}_1 - \hat{\mathbf{x}}_1)(\mathbf{x}_2 - \hat{\mathbf{x}}_2)^\dagger] \quad (65)$$

$$\mathbf{p}_2^* = \gamma_2^{-1} p_{2j}^* |h_{2j}|^2 \mathbb{E}_{\mathbf{y}}[(\mathbf{x}_2 - \hat{\mathbf{x}}_2)(\mathbf{x}_2 - \hat{\mathbf{x}}_2)^\dagger] - \gamma_2^{-1} p_{1j}^* |h_{2j}|^2 |h_{1j}| \mathbb{E}_{\mathbf{y}}[(\mathbf{x}_2 - \hat{\mathbf{x}}_2)(\mathbf{x}_1 - \hat{\mathbf{x}}_1)^\dagger] \quad (66)$$

In (65) and (66), due to the fact that the MSE is less than or equal to 1, and consider the special case where the second covariance term is zero, i.e., no user is interfering with the other, we can express the optimal power allocation with the single user mercury/waterfilling form [13] as follows:

$$\begin{cases} \mathbf{p}_1^* = \frac{1}{\text{snr}|h_{1j}|^2} \text{mmse}_1^{-1} \left(\frac{\gamma_1}{|h_{1j}|^2} \right) & \gamma_1 < |h_{1j}|^2 \\ \mathbf{p}_1^* = 0, & \gamma_1 \geq |h_{1j}|^2 \end{cases} \quad (67)$$

$$\begin{cases} \mathbf{p}_2^* = \frac{1}{\text{snr}|h_{2j}|^2} \text{mmse}_2^{-1} \left(\frac{\gamma_2}{|h_{2j}|^2} \right) & \gamma_2 < |h_{2j}|^2 \\ \mathbf{p}_2^* = 0, & \gamma_2 \geq |h_{2j}|^2 \end{cases} \quad (68)$$

There is a unique set \mathbf{p}_1^* , \mathbf{p}_2^* that satisfy the KKT conditions when the problem is strictly concave, corresponding to the global maximum.

VII. NUMERICAL RESULTS

In this section we will introduce a set of illustrative results of the two-user MAC with arbitrary inputs. In particular, we will present the simplified case where the two user channels are transmitting over scalar channels or single channel per user. We used Monte-Carlo method to generate the results.

Figure 1 illustrates the decay in the mutual information at the 45° when the two BPSK inputs cancel each other as they lie in the null space of the channel; called the Voronoi region, with the channels are such that $h_1 = 1$, and $h_2 = 1$.

If the channel gains or the total powers are different, the decay will be shifted into another line in the coordinate axis. However, we can easily check that if we induce orthogonality, for example if $h_1 = 1$ and $h_2 = 1j$, the mutual information will be the maximum achievable one without this decay region.

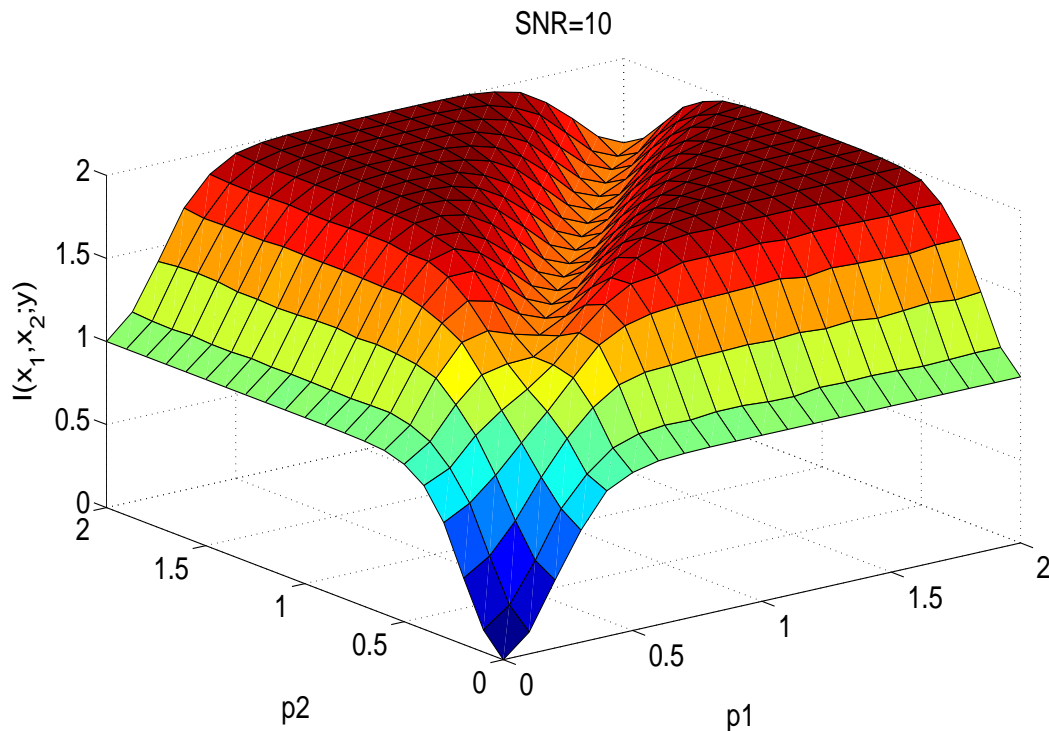


Figure 1. The two-user MAC mutual information with BPSK inputs.

In Figure 2 we show the total MMSE for the two-user MAC. It is interesting to notice that the MMSE behavior corresponds to swapping the mutual information curve in Figure 1. This can let us visualize the

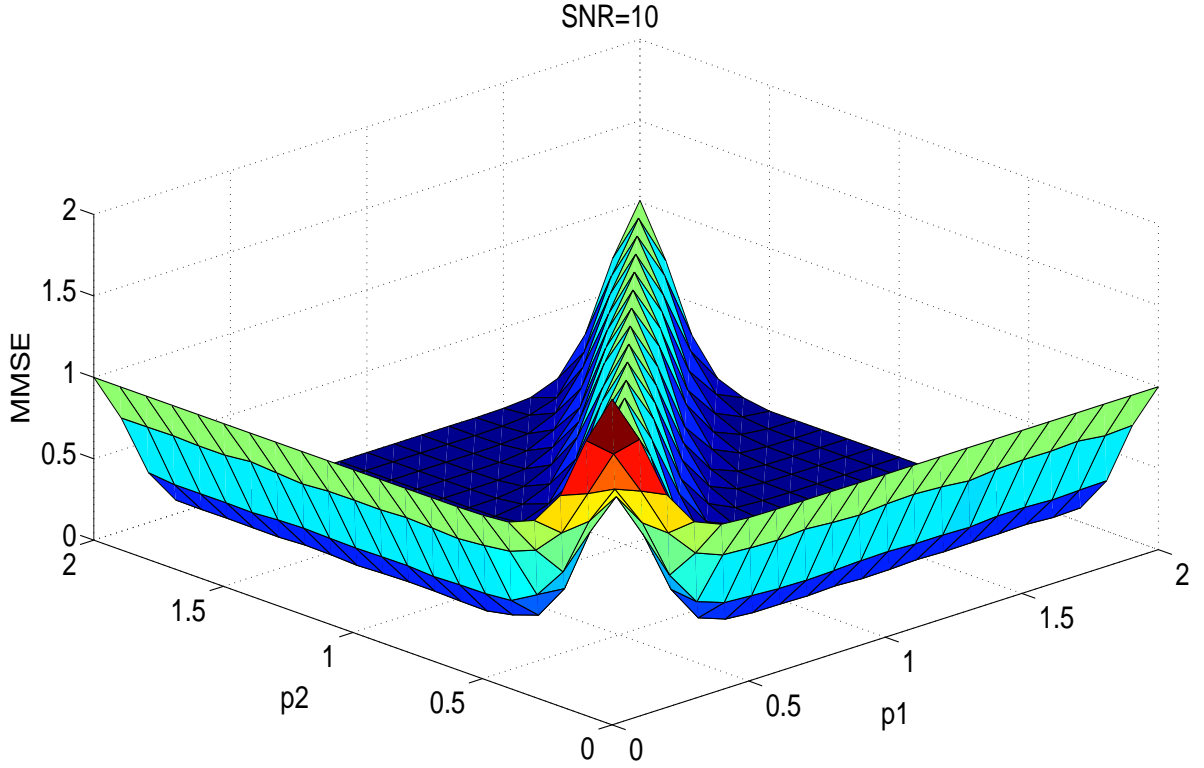


Figure 2. The two-user MAC MMSE with BPSK inputs.

relation between the gradient of the mutual information and the MMSE [10], [12].

Of particular relevance to observe that the decay in the mutual information saturates at 1.5 bits while the decay of the MMSE saturates at 1 for equal power of both user's, e.g., $(p_1, p_2) = (2, 2)$ in Figure 1, and Figure 2, respectively.

This is explained in the new fundamental relation which quantifies a loss of 0.5 bits in the data rates. For the case of Gaussian inputs the relation still holds, but the decay encountered in the arbitrary input setup will not exist. This can be easily verified theoretically and by simulation. For further analysis on Gaussian and mixed inputs, we refer the reader to [18].

Notice that the rate loss in the first user mutual information is due to the interference incurred by the second user. This can be understood via the negative second term in the gradient of the mutual information, similarly, for the other user. This term is a function of the main channel, the interferer channel, the main and interferer powers, and the input estimates.

Therefore, the interference from user 1 to user 2 can be interpreted as:

$$\mathbf{H}_1^\dagger \mathbf{H}_2 \mathbf{P}_2^* \mathbb{E}[\hat{\mathbf{x}}_2 \hat{\mathbf{x}}_1^\dagger], \quad (69)$$

where we can re-write equation (69) for the single channel per user case in terms of the covariance of the interferer as follows:

$$\frac{1}{snr \ h_1^\dagger h_2} cov(snr \ h_1^\dagger h_2 p_2^*), \quad (70)$$

Similarly, the interference from user 2 to user 1 can be interpreted as:

$$\mathbf{H}_2^\dagger \mathbf{H}_1 \mathbf{P}_1^* \mathbb{E}[\hat{\mathbf{x}}_1 \hat{\mathbf{x}}_2^\dagger], \quad (71)$$

Where we can re-write equation (71) for the single channel per user case in terms of the covariance of the interferer as follows:

$$\frac{1}{snr \ h_2^\dagger h_1} cov(snr \ h_2^\dagger h_1 p_1^*). \quad (72)$$

Figure 3 and Figure 4 illustrate the MMSE per user in the two-user MAC channel. MMSE1 and MMSE2 correspond to the mean-squared error of the first and second user respectively. It is worth to note that the sum of the two MMSEs leads to the total MMSE in Figure 2.

Figure 5 illustrates the covariance of one interferer into the other user. We can see the negative term that leads to the loss incurred in the achieved mutual information.

Finally, Figure 6 presents the optimal power allocation for the two-user MAC. When the total power for the first user is larger than that for the second user, the optimal power allocation for the first will be to allocate its total power, as far as the second user doesn't extremely interfere.

However, for the second user, the optimal power allocation will start with an allocation of the total power, then iteratively optimize it by decreasing it. Therefore, precoding is of great importance to such cases where we can align the transmit directions - or caused interference - and maximize the information rates.

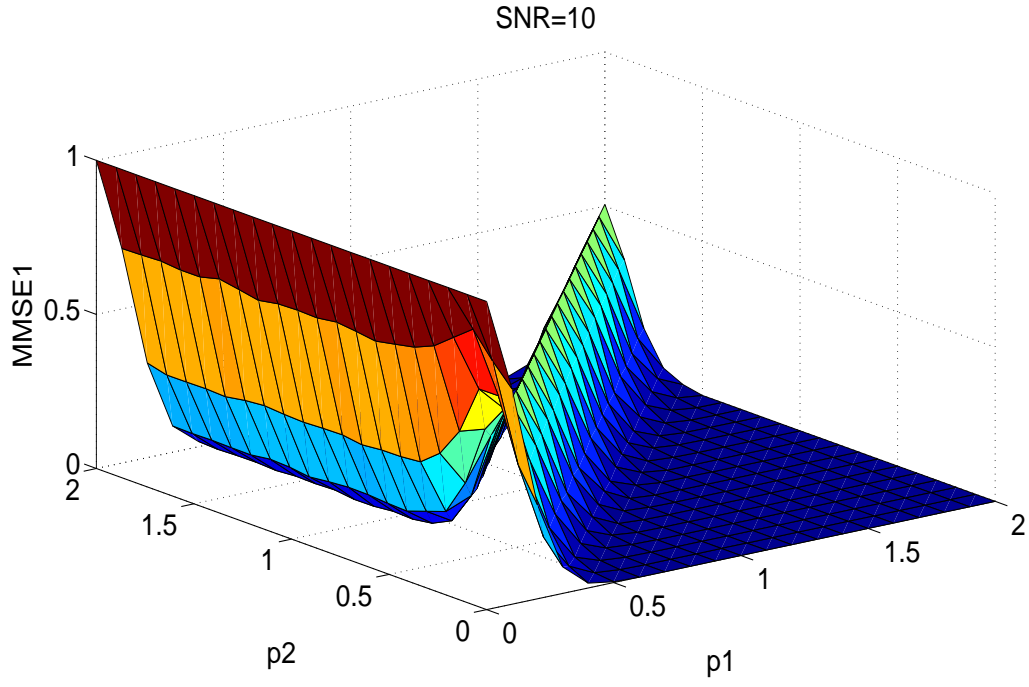


Figure 3. MMSE1 for the first user of the two-user MAC with BPSK input.

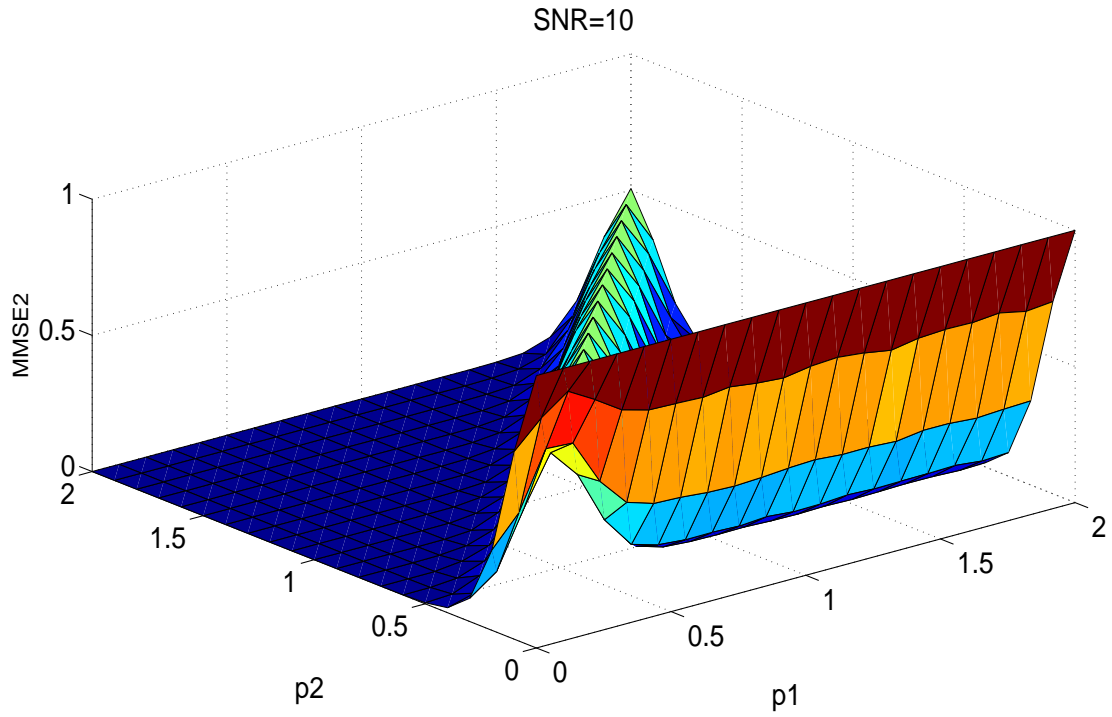


Figure 4. MMSE2 for the second user of the two-user MAC with BPSK input.

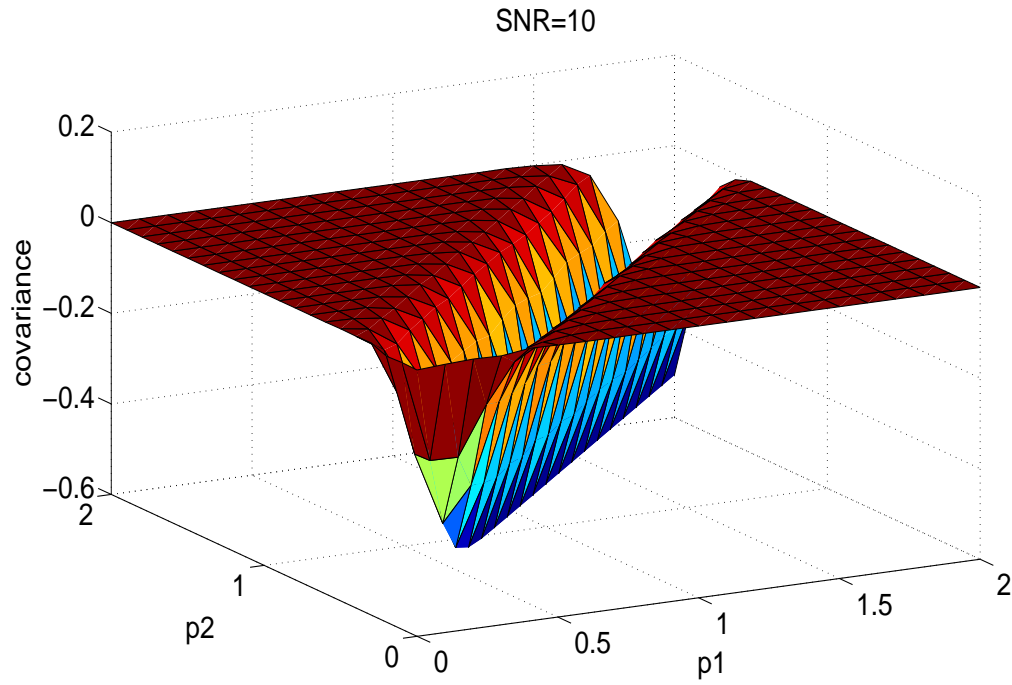


Figure 5. Covariance for the first user of the two-user MAC with BPSK input.

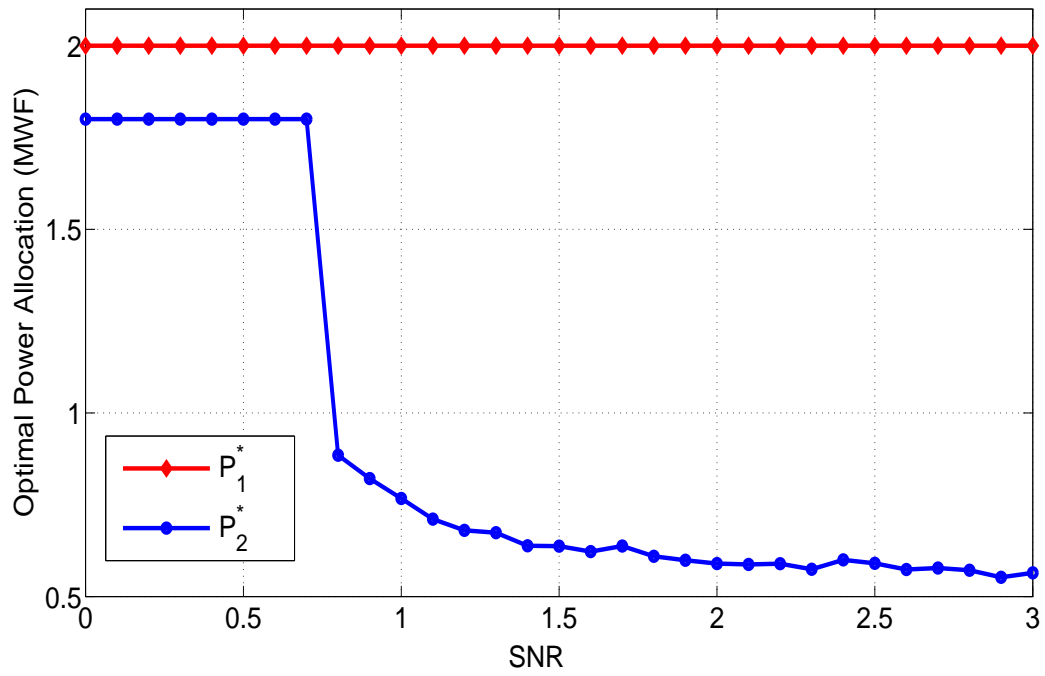


Figure 6. Optimal power allocation of the two user MAC with BPSK inputs.

VIII. CONCLUSIONS

In this paper, we studied the two-user MAC Gaussian channels with arbitrary inputs. We provide a generalization of the fundamental relation between the mutual information and the MMSE to a new fundamental relation which applies to multiuser channel setups. Further, we proved our generalized framework deriving the relation between the gradient of the mutual information and the MMSE for any input setup. We build upon derivations for the optimal precoding and optimal power allocation. We specialize our results to different cases of the input users, where for the simple case when each user transmits over a SISO channel, the relations can be mapped to the ones by Guo et al. for point to point channels with general distributions [12], and by Palomar et al. for the linear vector Gaussian channels with mutually interfering inputs [10]. We have shown that the optimal precoder can follow a fixed point equation similar to [3] under certain conditions, where we identified the optimal precoder structure with respect to the power allocation, and the right and left singular vectors. Much simplification to the setup where both inputs don't interfere with each other leads to the mercury/waterfilling interpretation for the single user case with binary inputs [13]. This casts further insights to have an interpretation of the interference as discussed. Additionally, it casts further research questions in quantifying the losses incurred in the mutual information due to interference with and without precoding, and in conjunction with error correcting codes. Furthermore, we have analyzed the system at the low-SNR regime, we have derived novel closed forms of the MMSE and the mutual information at the low SNR based on our new unveiled fundamental relation. The interference at the low-SNR adds non-linear terms at the low SNR which decides on how fast the capacity can be achieved with respect to the SNR.

Of particular relevance are the implications of our derived new fundamental relation in the study of the fundamental limits of cooperation, if a multi-cell processing framework is to be implemented between the BSs in a network. In particular, the interpretation of the interference in the studied two-user MAC, which is basically extracted from the gradient formula, can be extended to the k -user MAC channel where the impact of interference incurs higher losses into the information rates. Therefore, this has direct impact on

the optimal designs of power allocation and precoding. In fact, a global CSI and/or data sharing will not be possible, and this enforces a framework where multi-cell processing (MCP) should be implemented in clusters of known size. Furthermore, it is straightforward to think about setups where the cooperation and a clustered knowledge, i.e., a MCP with different levels of cooperation, can provide robust techniques to cancel or decode the interference at the transmitter/receiver sides and so improving the network spectral efficiency. In addition, its of particular relevance to understand the significance of our derivations on setups that include higher number of users accessing the network. For instance, the application of such formulas into scenarios where main users and cognitive users co-exist. The main users can design there optimal power allocation and optimal precoding subject to, the known interferer(s), i.e., the cognitive user interference, whether, they will be transmitting over SISO channels, or using OFDM modulation. Further, we have also she light on a new class of problems where a multi-tier precoding could exist, one example is massive MIMOs. In particular, a new formula can be utilized to express the signal to interference and noise ratio (SINR) based on the interpretation of the interference.

IX. APPENDIX A: PROOF OF THEOREM 1

The conditional probability density for the two-user MAC can be written as follows:

$$p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{\pi^{n_r}} e^{-\|\mathbf{y} - \sqrt{snr}\mathbf{H}_1\mathbf{P}_1\mathbf{x}_1 - \sqrt{snr}\mathbf{H}_2\mathbf{P}_2\mathbf{x}_2\|^2} \quad (73)$$

Thus, the corresponding mutual information is:

$$I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}) = \mathbb{E} \left[\log \left(\frac{p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2)}{p_y(\mathbf{y})} \right) \right] \quad (74)$$

$$I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}) = -n_r \log(\pi e) - \mathbb{E} [\log(p_y(\mathbf{y}))] \quad (75)$$

$$I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}) = -n_r \log(\pi e) - \int p_y(\mathbf{y}) \log(p_y(\mathbf{y})) d\mathbf{y} \quad (76)$$

Then, the derivative of the mutual information with respect to the SNR is as follows:

$$\frac{dI(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y})}{dsnr} = -\frac{\partial}{\partial snr} \int p_y(\mathbf{y}) \log(p_y(\mathbf{y})) d\mathbf{y} \quad (77)$$

$$= - \int \left(p_y(\mathbf{y}) \frac{1}{p_y(\mathbf{y})} + \log(p_y(\mathbf{y})) \right) \frac{\partial p_y(\mathbf{y})}{\partial snr} d\mathbf{y} \quad (78)$$

$$= - \int (1 + \log(p_y(\mathbf{y}))) \frac{\partial p_y(\mathbf{y})}{\partial snr} d\mathbf{y} \quad (79)$$

Where the probability density function of the received vector \mathbf{y} is given by:

$$p_y(\mathbf{y}) = \sum_{\mathbf{x}_1, \mathbf{x}_2} p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) p_{x_1, x_2}(\mathbf{x}_1, \mathbf{x}_2) \quad (80)$$

$$= \mathbb{E}_{x_1, x_2} [p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2)] \quad (81)$$

The derivative of the conditional output with respect to the SNR can be written as:

$$\begin{aligned} \frac{\partial p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2)}{\partial snr} = & \\ & - p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) \frac{\partial}{\partial snr} (\mathbf{y} - \sqrt{snr} \mathbf{H}_1 \mathbf{P}_1 \mathbf{x}_1 - \sqrt{snr} \mathbf{H}_2 \mathbf{P}_2 \mathbf{x}_2)^\dagger \times \\ & (\mathbf{y} - \sqrt{snr} \mathbf{H}_1 \mathbf{P}_1 \mathbf{x}_1 - \sqrt{snr} \mathbf{H}_2 \mathbf{P}_2 \mathbf{x}_2) \end{aligned} \quad (82)$$

$$\begin{aligned} = & - \frac{1}{\sqrt{snr}} ((\mathbf{H}_1 \mathbf{P}_1 \mathbf{x}_1)^\dagger - (\mathbf{H}_2 \mathbf{P}_2 \mathbf{x}_2)^\dagger) (\mathbf{y} - \sqrt{snr} \mathbf{H}_1 \mathbf{P}_1 \mathbf{x}_1 - \sqrt{snr} \mathbf{H}_2 \mathbf{P}_2 \mathbf{x}_2) \times \\ & p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) \end{aligned} \quad (83)$$

$$= - \frac{1}{\sqrt{snr}} ((\mathbf{H}_1 \mathbf{P}_1 \mathbf{x}_1)^\dagger - (\mathbf{H}_2 \mathbf{P}_2 \mathbf{x}_2)^\dagger) \nabla_{\mathbf{y}} p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) \quad (84)$$

Therefore, we have:

$$\begin{aligned} \mathbb{E}_{x_1, x_2} [\nabla_{snr} p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2)] = & \\ & \mathbb{E}_{x_1, x_2} \left[- \frac{1}{\sqrt{snr}} ((\mathbf{H}_1 \mathbf{P}_1 \mathbf{x}_1)^\dagger - (\mathbf{H}_2 \mathbf{P}_2 \mathbf{x}_2)^\dagger) \nabla_{\mathbf{y}} p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) \right] \end{aligned} \quad (85)$$

Substitute (85) into (79), we get:

$$\begin{aligned} \frac{dI(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y})}{dsnr} = & \frac{1}{\sqrt{snr}} \int (1 + \log(p_y(\mathbf{y}))) \mathbb{E}_{x_1, x_2} [((\mathbf{H}_1 \mathbf{P}_1 \mathbf{x}_1)^\dagger - (\mathbf{H}_2 \mathbf{P}_2 \mathbf{x}_2)^\dagger) \times \\ & \nabla_{\mathbf{y}} p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2)] d\mathbf{y} \end{aligned} \quad (86)$$

$$= \frac{1}{\sqrt{snr}} \mathbb{E}_{x_1, x_2} \left[\left(\int (1 + \log(p_y(\mathbf{y}))) ((\mathbf{H}_1 \mathbf{P}_1 \mathbf{x}_1)^\dagger - (\mathbf{H}_2 \mathbf{P}_2 \mathbf{x}_2)^\dagger) \nabla_{\mathbf{y}} p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) d\mathbf{y} \right) \right] \quad (87)$$

Using integration by parts applied to the real and imaginary parts of \mathbf{y} we have:

$$\int (1 + \log(p_y(\mathbf{y}))) \frac{\partial p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2)}{\partial t} d\mathbf{t} = \int (1 + \log(p_y(\mathbf{y}))) p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \frac{1}{p_y(\mathbf{y})} \frac{\partial p_y(\mathbf{y})}{\partial t} p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) d\mathbf{t} \quad (88)$$

The first term in (88) goes to zero as $\|\mathbf{y}\| \rightarrow \infty$. Therefore,

$$\frac{dI(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y})}{dsnr} = \frac{1}{\sqrt{snr}} \mathbb{E}_{x_1, x_2} \left[- \int \left(((\mathbf{H}_1 \mathbf{P}_1 \mathbf{x}_1)^\dagger - (\mathbf{H}_2 \mathbf{P}_2 \mathbf{x}_2)^\dagger) \frac{p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2)}{p_y(\mathbf{y})} \nabla_{\mathbf{y}} p_y(\mathbf{y}) d\mathbf{y} \right) \right] \quad (89)$$

$$\frac{dI(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y})}{dsnr} = - \frac{1}{\sqrt{snr}} \int \nabla_{\mathbf{y}} p_y(\mathbf{y}) \mathbb{E}_{x_1, x_2} \left[((\mathbf{H}_1 \mathbf{P}_1 \mathbf{x}_1)^\dagger - (\mathbf{H}_2 \mathbf{P}_2 \mathbf{x}_2)^\dagger) \frac{p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2)}{p_y(\mathbf{y})} \right] d\mathbf{y} \quad (90)$$

$$\frac{dI(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y})}{dsnr} = - \frac{1}{\sqrt{snr}} \int \nabla_{\mathbf{y}} p_y(\mathbf{y}) \mathbb{E}_{x_1, x_2} \left((\mathbf{H}_1 \mathbf{P}_1)^\dagger \mathbb{E}_{x_1|y} [\mathbf{x}_1|\mathbf{y}]^\dagger - (\mathbf{H}_2 \mathbf{P}_2)^\dagger \mathbb{E}_{x_2|y} [\mathbf{x}_2|\mathbf{y}]^\dagger \right) d\mathbf{y} \quad (91)$$

However,

$$\begin{aligned} \nabla_{\mathbf{y}} p_y(\mathbf{y}) &= \nabla_{\mathbf{y}} \mathbb{E}_{x_1, x_2} [p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2)] \\ &= \mathbb{E}_{x_1, x_2} [\nabla_{\mathbf{y}} p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2)] \\ &= -\mathbb{E}_{x_1, x_2} [p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) (\mathbf{y} - \sqrt{snr} \mathbf{H}_1 \mathbf{P}_1 \mathbf{x}_1 - \sqrt{snr} \mathbf{H}_2 \mathbf{P}_2 \mathbf{x}_2)] \\ &= -\mathbb{E}_{x_1, x_2} [p_y(\mathbf{y}) (\mathbf{y} - \sqrt{snr} \mathbf{H}_1 \mathbf{P}_1 \mathbf{x}_1 - \sqrt{snr} \mathbf{H}_2 \mathbf{P}_2 \mathbf{x}_2) | \mathbf{y}] \\ &= -p_y(\mathbf{y}) (\mathbf{y} - \sqrt{snr} \mathbf{H}_1 \mathbf{P}_1 \mathbb{E}_{x_1|y} [\mathbf{x}_1|\mathbf{y}] - \sqrt{snr} \mathbf{H}_2 \mathbf{P}_2 \mathbb{E}_{x_2|y} [\mathbf{x}_2|\mathbf{y}]) \end{aligned} \quad (92)$$

Substitute (92) into (91) we get:

$$\begin{aligned} \frac{dI(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y})}{dsnr} &= \frac{1}{\sqrt{snr}} \int p_y(\mathbf{y}) (\mathbf{y} - \sqrt{snr} \mathbf{H}_1 \mathbf{P}_1 \mathbb{E}_{x_1|y} [\mathbf{x}_1|\mathbf{y}] - \sqrt{snr} \mathbf{H}_2 \mathbf{P}_2 \mathbb{E}_{x_2|y} [\mathbf{x}_2|\mathbf{y}]) \times \\ &\quad \mathbb{E}_{x_1, x_2} \left((\mathbf{H}_1 \mathbf{P}_1)^\dagger \mathbb{E}_{x_1|y} [\mathbf{x}_1|\mathbf{y}]^\dagger - (\mathbf{H}_2 \mathbf{P}_2)^\dagger \mathbb{E}_{x_2|y} [\mathbf{x}_2|\mathbf{y}]^\dagger \right) d\mathbf{y} \end{aligned} \quad (93)$$

$$\begin{aligned} \frac{dI(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y})}{dsnr} &= \frac{1}{\sqrt{snr}} \mathbb{E}_{\mathbf{y}} [\mathbf{y} \mathbf{x}_1^\dagger] (\mathbf{H}_1 \mathbf{P}_1)^\dagger - \frac{1}{\sqrt{snr}} \mathbb{E}_{\mathbf{y}} [\mathbf{y} \mathbf{x}_2^\dagger] (\mathbf{H}_2 \mathbf{P}_2)^\dagger \\ &\quad - \mathbb{E}_{\mathbf{y}} [\mathbf{H}_1 \mathbf{P}_1 \mathbb{E}_{x_1|y} [\mathbf{x}_1|\mathbf{y}] \mathbb{E}_{x_1|y} [\mathbf{x}_1|\mathbf{y}]^\dagger (\mathbf{H}_1 \mathbf{P}_1)^\dagger + \mathbb{E}_{\mathbf{y}} [\mathbf{H}_1 \mathbf{P}_1 \mathbb{E}_{x_1|y} [\mathbf{x}_1|\mathbf{y}] \mathbb{E}_{x_2|y} [\mathbf{x}_2|\mathbf{y}]^\dagger] (\mathbf{H}_2 \mathbf{P}_2)^\dagger \\ &\quad + \mathbb{E}_{\mathbf{y}} [\mathbf{H}_2 \mathbf{P}_2 \mathbb{E}_{x_2|y} [\mathbf{x}_2|\mathbf{y}] \mathbb{E}_{x_2|y} [\mathbf{x}_2|\mathbf{y}]^\dagger (\mathbf{H}_2 \mathbf{P}_2)^\dagger - \mathbb{E}_{\mathbf{y}} [\mathbf{H}_2 \mathbf{P}_2 \mathbb{E}_{x_2|y} [\mathbf{x}_2|\mathbf{y}] \mathbb{E}_{x_1|y} [\mathbf{x}_1|\mathbf{y}]^\dagger] (\mathbf{H}_1 \mathbf{P}_1)^\dagger \end{aligned} \quad (94)$$

Therefore,

$$\begin{aligned}
\frac{dI(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y})}{dsnr} &= \mathbf{H}_1 \mathbf{P}_1 \mathbb{E}_{x_1}[\mathbf{x}_1 \mathbf{x}_1^\dagger] (\mathbf{H}_1 \mathbf{P}_1)^\dagger \\
&\quad - \mathbf{H}_1 \mathbf{P}_1 \mathbb{E}_y[\mathbb{E}_{x_1|y}[\mathbf{x}_1 | \mathbf{y}] \mathbb{E}_{x_1|y}[\mathbf{x}_1 | \mathbf{y}]^\dagger] (\mathbf{H}_1 \mathbf{P}_1)^\dagger \\
&\quad + \mathbf{H}_1 \mathbf{P}_1 \mathbb{E}_y[\mathbb{E}_{x_1|y}[\mathbf{x}_1 | \mathbf{y}] \mathbb{E}_{x_2|y}[\mathbf{x}_2 | \mathbf{y}]^\dagger] (\mathbf{H}_2 \mathbf{P}_2)^\dagger \\
&\quad - \mathbf{H}_2 \mathbf{P}_2 \mathbb{E}_{x_2}[\mathbf{x}_2 \mathbf{x}_2^\dagger] (\mathbf{H}_2 \mathbf{P}_2)^\dagger \\
&\quad + \mathbf{H}_2 \mathbf{P}_2 \mathbb{E}_y[\mathbb{E}_{x_2|y}[\mathbf{x}_2 | \mathbf{y}] \mathbb{E}_{x_2|y}[\mathbf{x}_2 | \mathbf{y}]^\dagger] (\mathbf{H}_2 \mathbf{P}_2)^\dagger \\
&\quad - \mathbf{H}_2 \mathbf{P}_2 \mathbb{E}_y[\mathbb{E}_{x_1|y}[\mathbf{x}_1 | \mathbf{y}] \mathbb{E}_{x_2|y}[\mathbf{x}_2 | \mathbf{y}]^\dagger] (\mathbf{H}_1 \mathbf{P}_1)^\dagger \\
\\
\frac{dI(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y})}{dsnr} &= \mathbf{H}_1 \mathbf{P}_1 \mathbf{E}_1 (\mathbf{H}_1 \mathbf{P}_1)^\dagger - \mathbf{H}_2 \mathbf{P}_2 \mathbf{E}_2 (\mathbf{H}_2 \mathbf{P}_2)^\dagger \\
&\quad + \mathbf{H}_1 \mathbf{P}_1 \mathbb{E}_y[\mathbb{E}_{x_1|y}[\mathbf{x}_1 | \mathbf{y}] \mathbb{E}_{x_2|y}[\mathbf{x}_2 | \mathbf{y}]^\dagger] (\mathbf{H}_2 \mathbf{P}_2)^\dagger \\
&\quad - \mathbf{H}_2 \mathbf{P}_2 \mathbb{E}_y[\mathbb{E}_{x_2|y}[\mathbf{x}_2 | \mathbf{y}] \mathbb{E}_{x_1|y}[\mathbf{x}_1 | \mathbf{y}]^\dagger] (\mathbf{H}_1 \mathbf{P}_1)^\dagger
\end{aligned}$$

Therefore, due to the fact that the expectation remains the same if $\mathbf{y} \sim \mathcal{CN}(\sqrt{snr} \mathbf{H}_1 \mathbf{P}_1 \mathbf{x}_1 + \sqrt{snr} \mathbf{H}_2 \mathbf{P}_2 \mathbf{x}_2, \mathbf{I})$ or $\mathbf{y} \sim \mathcal{CN}(-\sqrt{snr} \mathbf{H}_1 \mathbf{P}_1 \mathbf{x}_1 - \sqrt{snr} \mathbf{H}_2 \mathbf{P}_2 \mathbf{x}_2, \mathbf{I})$, therefore, $\mathbb{E}[\mathbf{x}_1 \mathbf{x}_1^\dagger] = -\mathbb{E}[\mathbf{x}_1 \mathbf{x}_1^\dagger]$ and $\mathbb{E}[\mathbb{E}[\mathbf{x}_1 | \mathbf{y}] \mathbb{E}[\mathbf{x}_1 | \mathbf{y}]^\dagger] = -\mathbb{E}[\mathbb{E}[\mathbf{x}_1 | \mathbf{y}] \mathbb{E}[\mathbf{x}_1 | \mathbf{y}]^\dagger]$, and similarly for the other input, the derivative of the mutual information with respect to the SNR and the per users mmse and input estimates (or covariances) is as follows:

$$\begin{aligned}
\frac{dI(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y})}{dsnr} &= mmse_1(snr) + mmse_2(snr) + Tr \left\{ \mathbf{H}_1 \mathbf{P}_1 \mathbb{E}_y[\hat{\mathbf{x}}_1 \hat{\mathbf{x}}_2^\dagger] (\mathbf{H}_2 \mathbf{P}_2)^\dagger \right\} \\
&\quad - Tr \left\{ \mathbf{H}_2 \mathbf{P}_2 \mathbb{E}_y[\hat{\mathbf{x}}_2 \hat{\mathbf{x}}_1^\dagger] (\mathbf{H}_1 \mathbf{P}_1)^\dagger \right\}
\end{aligned}$$

Therefore, we can write the derivative of the derivative of the mutual information with respect to the snr as follows:

$$\frac{dI(snr)}{dsnr} = mmse(snr) + \psi(snr) \tag{95}$$

Therefore, Theorem 1 has been proved as a generalization of the one by Guo, Shamai, Verdu in [12] to the multiuser case.

X. APPENDIX B: PROOF OF THEOREM 3

The conditional probability density for the two-user MAC can be written as follows:

$$p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{\pi^{n_r}} e^{-\|y - \sqrt{snr}\mathbf{H}_1\mathbf{P}_1\mathbf{x}_1 - \sqrt{snr}\mathbf{H}_2\mathbf{P}_2\mathbf{x}_2\|^2} \quad (96)$$

Thus, the corresponding mutual information is:

$$I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}) = \mathbb{E} \left[\log \left(\frac{p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2)}{p_y(\mathbf{y})} \right) \right] \quad (97)$$

$$I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}) = -n_r \log(\pi e) - \mathbb{E} [\log(p_y(\mathbf{y}))] \quad (98)$$

Then, the gradient of the mutual information with respect to the channel of user 1 on the two-user MAC is as follows:

$$\frac{\partial I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y})}{\partial \mathbf{H}_1^\dagger} = -\frac{\partial}{\partial \mathbf{H}_1^\dagger} \int p_y(\mathbf{y}) \log(p_y(\mathbf{y})) d\mathbf{y} \quad (99)$$

$$= -\int \left(p_y(\mathbf{y}) \frac{1}{p_y(\mathbf{y})} + \log(p_y(\mathbf{y})) \right) \frac{\partial p_y(\mathbf{y})}{\partial \mathbf{H}_1^\dagger} d\mathbf{y} \quad (100)$$

$$= -\int (1 + \log(p_y(\mathbf{y}))) \frac{\partial p_y(\mathbf{y})}{\partial \mathbf{H}_1^\dagger} d\mathbf{y} \quad (101)$$

Where the probability density function of the received vector \mathbf{y} is given by:

$$p_y(\mathbf{y}) = \sum_{\mathbf{x}_1, \mathbf{x}_2} p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) p_{x_1, x_2}(\mathbf{x}_1, \mathbf{x}_2) \quad (102)$$

$$= \mathbb{E}_{x_1, x_2} [p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2)] \quad (103)$$

The derivative of the conditional output can be written as:

$$\begin{aligned} \frac{\partial p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2)}{\partial \mathbf{H}_1^\dagger} &= \\ &- p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) \frac{\partial}{\partial \mathbf{H}_1^\dagger} (\mathbf{y} - \sqrt{snr}\mathbf{H}_1\mathbf{P}_1\mathbf{x}_1 - \sqrt{snr}\mathbf{H}_2\mathbf{P}_2\mathbf{x}_2)^\dagger \times \\ &\quad (\mathbf{y} - \sqrt{snr}\mathbf{H}_1\mathbf{P}_1\mathbf{x}_1 - \sqrt{snr}\mathbf{H}_2\mathbf{P}_2\mathbf{x}_2) \end{aligned} \quad (104)$$

$$= p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) (\mathbf{y} - \sqrt{snr}\mathbf{H}_1\mathbf{P}_1\mathbf{x}_1 - \sqrt{snr}\mathbf{H}_2\mathbf{P}_2\mathbf{x}_2) \sqrt{snr}\mathbf{x}_1^\dagger \mathbf{P}_1^\dagger \quad (105)$$

$$= -\nabla_{\mathbf{y}} p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) \sqrt{snr} \mathbf{x}_1^\dagger \mathbf{P}_1^\dagger \quad (106)$$

Therefore, we have:

$$\mathbb{E}_{x_1, x_2} \left[\nabla_{\mathbf{H}_1^\dagger} p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) \right] = \mathbb{E}_{x_1, x_2} \left[-\nabla_{\mathbf{y}} p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) \sqrt{snr} \mathbf{x}_1^\dagger \mathbf{P}_1^\dagger \right] \quad (107)$$

Substitute (107) into (101), we get:

$$\frac{\partial I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y})}{\partial \mathbf{H}_1^\dagger} = \int (1 + \log(p_y(\mathbf{y}))) \mathbb{E}_{x_1, x_2} [\nabla_{\mathbf{y}} p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) \times \sqrt{snr} \mathbf{x}_1^\dagger \mathbf{P}_1^\dagger] d\mathbf{y} \quad (108)$$

$$= \mathbb{E}_{x_1, x_2} \left[\left(\int (1 + \log(p_y(\mathbf{y}))) \nabla_{\mathbf{y}} p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) d\mathbf{y} \right) \sqrt{snr} \mathbf{x}_1^\dagger \mathbf{P}_1^\dagger \right] \quad (109)$$

Using integration by parts applied to the real and imaginary parts of \mathbf{y} we have:

$$\int (1 + \log(p_y(\mathbf{y}))) \frac{\partial p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2)}{\partial t} d\mathbf{t} = \int (1 + \log(p_y(\mathbf{y}))) p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \frac{1}{p_y(\mathbf{y})} \frac{\partial p_y(\mathbf{y})}{\partial t} p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) d\mathbf{t} \quad (110)$$

The first term in (110) goes to zero as $\|\mathbf{y}\| \rightarrow \infty$. Therefore,

$$\frac{\partial I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y})}{\partial \mathbf{H}_1^\dagger} = \mathbb{E}_{x_1, x_2} \left[- \int \left(\frac{p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2)}{p_y(\mathbf{y})} \nabla_{\mathbf{y}} p_y(\mathbf{y}) d\mathbf{y} \right) \sqrt{snr} \mathbf{x}_1^\dagger \mathbf{P}_1^\dagger \right] \quad (111)$$

$$\frac{\partial I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y})}{\partial \mathbf{H}_1^\dagger} = - \int \nabla_{\mathbf{y}} p_y(\mathbf{y}) \mathbb{E}_{x_1, x_2} \left[\frac{p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2)}{p_y(\mathbf{y})} \sqrt{snr} \mathbf{x}_1^\dagger \mathbf{P}_1^\dagger \right] d\mathbf{y} \quad (112)$$

$$\frac{\partial I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y})}{\partial \mathbf{H}_1^\dagger} = -\sqrt{snr} \int \nabla_{\mathbf{y}} p_y(\mathbf{y}) \mathbb{E}_{x_1|y} [\mathbf{x}_1 | \mathbf{y}]^\dagger \mathbf{P}_1^\dagger d\mathbf{y} \quad (113)$$

However,

$$\begin{aligned} \nabla_{\mathbf{y}} p_y(\mathbf{y}) &= \nabla_{\mathbf{y}} \mathbb{E}_{x_1, x_2} [p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2)] \\ &= \mathbb{E}_{x_1, x_2} [\nabla_{\mathbf{y}} p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2)] \\ &= -\mathbb{E}_{x_1, x_2} [p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) (\mathbf{y} - \sqrt{snr} \mathbf{H}_1 \mathbf{P}_1 \mathbf{x}_1 - \sqrt{snr} \mathbf{H}_2 \mathbf{P}_2 \mathbf{x}_2)] \\ &= -\mathbb{E}_{x_1, x_2} [p_y(\mathbf{y}) (\mathbf{y} - \sqrt{snr} \mathbf{H}_1 \mathbf{P}_1 \mathbf{x}_1 - \sqrt{snr} \mathbf{H}_2 \mathbf{P}_2 \mathbf{x}_2) | \mathbf{y}] \\ &= -p_y(\mathbf{y}) (\mathbf{y} - \sqrt{snr} \mathbf{H}_1 \mathbf{P}_1 \mathbb{E}_{\mathbf{x}_1|y} [\mathbf{x}_1 | \mathbf{y}] - \sqrt{snr} \mathbf{H}_2 \mathbf{P}_2 \mathbb{E}_{\mathbf{x}_2|y} [\mathbf{x}_2 | \mathbf{y}]) \end{aligned} \quad (114)$$

Substitute (114) into (113) we get:

$$\frac{\partial I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y})}{\partial \mathbf{H}_1^\dagger} = \int p_y(\mathbf{y}) (\mathbf{y} - \sqrt{snr} \mathbf{H}_1 \mathbf{P}_1 \mathbb{E}_{x_1|y}[\mathbf{x}_1|\mathbf{y}] - \sqrt{snr} \mathbf{H}_2 \mathbf{P}_2 \mathbb{E}_{x_2|y}[\mathbf{x}_2|\mathbf{y}]) \times \sqrt{snr} \mathbb{E}_{x_1|y}[\mathbf{x}_1|\mathbf{y}]^\dagger \mathbf{P}_1^\dagger d\mathbf{y} \quad (115)$$

$$\begin{aligned} \frac{\partial I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y})}{\partial \mathbf{H}_1^\dagger} &= \sqrt{snr} \mathbb{E}_y[\mathbf{y} \mathbf{x}_1^\dagger] \mathbf{P}_1^\dagger \\ &\quad - snr \mathbb{E}_y[\mathbf{H}_1 \mathbf{P}_1 \mathbb{E}_{x_1|y}[\mathbf{x}_1|\mathbf{y}] \mathbb{E}_{x_1|y}[\mathbf{x}_1|\mathbf{y}]^\dagger] \mathbf{P}_1^\dagger \\ &\quad - snr \mathbb{E}_y[\mathbf{H}_2 \mathbf{P}_2 \mathbb{E}_{x_2|y}[\mathbf{x}_2|\mathbf{y}] \mathbb{E}_{x_2|y}[\mathbf{x}_1|\mathbf{y}]^\dagger] \mathbf{P}_1^\dagger \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{\partial I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y})}{\partial \mathbf{H}_1^\dagger} &= \sqrt{snr} \mathbf{H}_1 \mathbf{P}_1 \mathbb{E}_{x_1}[\mathbf{x}_1 \mathbf{x}_1^\dagger] \mathbf{P}_1^\dagger \\ &\quad - snr \mathbf{H}_1 \mathbf{P}_1 \mathbb{E}_y[\mathbb{E}_{x_1|y}[\mathbf{x}_1|\mathbf{y}] \mathbb{E}_{x_1|y}[\mathbf{x}_1|\mathbf{y}]^\dagger] \mathbf{P}_1^\dagger \\ &\quad - snr \mathbf{H}_2 \mathbf{P}_2 \mathbb{E}_y[\mathbb{E}_{x_2|y}[\mathbf{x}_2|\mathbf{y}] \mathbb{E}_{x_1|y}[\mathbf{x}_1|\mathbf{y}]^\dagger] \mathbf{P}_1^\dagger \\ \frac{\partial I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y})}{\partial \mathbf{H}_1^\dagger} &= snr \mathbf{H}_1 \mathbf{P}_1 \mathbf{E}_1 \mathbf{P}_1^\dagger - snr \mathbf{H}_2 \mathbf{P}_2 \mathbb{E}_y[\mathbb{E}_{x_2|y}[\mathbf{x}_2|\mathbf{y}] \mathbb{E}_{x_1|y}[\mathbf{x}_1|\mathbf{y}]^\dagger] \mathbf{P}_1^\dagger \end{aligned} \quad (116)$$

Similarly, we can derive the gradient of the mutual information in terms of the channel matrix that corresponds to the second user in the two-user MAC.

$$\frac{\partial I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y})}{\partial \mathbf{H}_2^\dagger} = snr \mathbf{H}_2 \mathbf{P}_2 \mathbf{E}_2 \mathbf{P}_2^\dagger - snr \mathbf{H}_1 \mathbf{P}_1 \mathbb{E}_y[\mathbb{E}_{x_1|y}[\mathbf{x}_1|\mathbf{y}] \mathbb{E}_{x_2|y}[\mathbf{x}_2|\mathbf{y}]^\dagger] \mathbf{P}_2^\dagger \quad (117)$$

Therefore, the gradient of the mutual information with respect to per user channel and the per user MMSE and input estimates (or covariances) is as follows:

$$\nabla_{\mathbf{H}_1} I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}) = snr \mathbf{H}_1 \mathbf{P}_1 \mathbf{E}_1 \mathbf{P}_1^\dagger - snr \mathbf{H}_2 \mathbf{P}_2 \mathbb{E}[\widehat{\mathbf{x}}_2 \widehat{\mathbf{x}}_1^\dagger] \mathbf{P}_1^\dagger \quad (118)$$

$$\nabla_{\mathbf{H}_2} I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}) = snr \mathbf{H}_2 \mathbf{P}_2 \mathbf{E}_2 \mathbf{P}_2^\dagger - snr \mathbf{H}_1 \mathbf{P}_1 \mathbb{E}[\widehat{\mathbf{x}}_1 \widehat{\mathbf{x}}_2^\dagger] \mathbf{P}_2^\dagger \quad (119)$$

Therefore, Theorem 3 has been proved.

XI. APPENDIX C: PROOF OF THEOREM 4

The gradient of the mutual information with respect to the precoding matrix of user 1 on the two-user MAC is as follows:

$$\frac{\partial I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y})}{\partial \mathbf{P}_1^\dagger} = -\frac{\partial}{\partial \mathbf{P}_1^\dagger} \int p_y(\mathbf{y}) \log(p_y(\mathbf{y})) d\mathbf{y} \quad (120)$$

$$= -\int \left(p_y(\mathbf{y}) \frac{1}{p_y(\mathbf{y})} + \log(p_y(\mathbf{y})) \right) \frac{\partial p_y(\mathbf{y})}{\partial \mathbf{P}_1^\dagger} d\mathbf{y} \quad (121)$$

$$= -\int (1 + \log(p_y(\mathbf{y}))) \frac{\partial p_y(\mathbf{y})}{\partial \mathbf{P}_1^\dagger} d\mathbf{y} \quad (122)$$

The derivative of the conditional output can be written as:

$$\begin{aligned} \frac{\partial p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2)}{\partial \mathbf{P}_1^\dagger} &= \\ &- p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) \frac{\partial}{\partial \mathbf{P}_1^\dagger} (\mathbf{y} - \sqrt{snr} \mathbf{H}_1 \mathbf{P}_1 \mathbf{x}_1 - \sqrt{snr} \mathbf{H}_2 \mathbf{P}_2 \mathbf{x}_2)^\dagger \times \\ &\quad (\mathbf{y} - \sqrt{snr} \mathbf{H}_1 \mathbf{P}_1 \mathbf{x}_1 - \sqrt{snr} \mathbf{H}_2 \mathbf{P}_2 \mathbf{x}_2) \end{aligned} \quad (123)$$

$$= p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) \mathbf{H}_1^\dagger (\mathbf{y} - \sqrt{snr} \mathbf{H}_1 \mathbf{P}_1 \mathbf{x}_1 - \sqrt{snr} \mathbf{H}_2 \mathbf{P}_2 \mathbf{x}_2) \mathbf{x}_1^\dagger \quad (124)$$

$$= -\sqrt{snr} \mathbf{H}_1^\dagger \nabla_{\mathbf{y}} p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) \mathbf{x}_1^\dagger \quad (125)$$

Therefore, we have:

$$\mathbb{E}_{x_1, x_2} \left[\nabla_{\mathbf{P}_1^\dagger} p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) \right] = \mathbb{E}_{x_1, x_2} \left[-\sqrt{snr} \mathbf{H}_1^\dagger \nabla_{\mathbf{y}} p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) \mathbf{x}_1^\dagger \right] \quad (126)$$

Substitute (126) into (122), we get:

$$\frac{\partial I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y})}{\partial \mathbf{P}_1^\dagger} = \int (1 + \log(p_y(\mathbf{y}))) \mathbb{E}_{x_1, x_2} \left[\sqrt{snr} \mathbf{H}_1^\dagger \nabla_{\mathbf{y}} p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) \mathbf{x}_1^\dagger \right] d\mathbf{y} \quad (127)$$

$$= \mathbb{E}_{x_1, x_2} \left[\sqrt{snr} \mathbf{H}_1^\dagger \left(\int (1 + \log(p_y(\mathbf{y}))) \nabla_{\mathbf{y}} p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) d\mathbf{y} \right) \mathbf{x}_1^\dagger \right] \quad (128)$$

Repeating the same steps as in (110), we have:

$$\frac{\partial I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y})}{\partial \mathbf{P}_1^\dagger} = \mathbb{E}_{x_1, x_2} \left[-\int \sqrt{snr} \mathbf{H}_1^\dagger \left(\frac{p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2)}{p_y(\mathbf{y})} \nabla_{\mathbf{y}} p_y(\mathbf{y}) d\mathbf{y} \right) \mathbf{x}_1^\dagger \right] \quad (129)$$

$$\frac{\partial I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y})}{\partial \mathbf{P}_1^\dagger} = -\int \sqrt{snr} \mathbf{H}_1^\dagger \nabla_{\mathbf{y}} p_y(\mathbf{y}) \mathbb{E}_{x_1, x_2} \left[\frac{p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2)}{p_y(\mathbf{y})} \mathbf{x}_1^\dagger \right] d\mathbf{y} \quad (130)$$

$$\frac{\partial I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y})}{\partial \mathbf{P}_1^\dagger} = -\sqrt{snr} \int \mathbf{H}_1^\dagger \nabla_{\mathbf{y}} p_y(\mathbf{y}) \mathbb{E}_{x_1|y} [\mathbf{x}_1 | \mathbf{y}]^\dagger d\mathbf{y} \quad (131)$$

Substitute (114) into (131) we get:

$$\begin{aligned} \frac{\partial I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y})}{\partial \mathbf{P}_1^\dagger} &= \int \sqrt{snr} \mathbf{H}_1^\dagger p_y(\mathbf{y}) \times \\ &\quad (\mathbf{y} - \sqrt{snr} \mathbf{H}_1 \mathbf{P}_1 \mathbb{E}_{x_1|y} [\mathbf{x}_1 | \mathbf{y}] - \sqrt{snr} \mathbf{H}_2 \mathbf{P}_2 \mathbb{E}_{x_2|y} [\mathbf{x}_2 | \mathbf{y}]) \mathbb{E}_{x_1|y} [\mathbf{x}_1 | \mathbf{y}]^\dagger d\mathbf{y} \end{aligned}$$

$$\begin{aligned} \frac{\partial I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y})}{\partial \mathbf{P}_1^\dagger} &= \sqrt{snr} \mathbf{H}_1^\dagger \mathbb{E}_y [\mathbf{y} \mathbf{x}_1^\dagger] - snr \mathbf{H}_1^\dagger \mathbb{E}_y [\mathbf{H}_1 \mathbf{P}_1 \mathbb{E}_{x_1|y} [\mathbf{x}_1 | \mathbf{y}] \mathbb{E}_{x_1|y} [\mathbf{x}_1 | \mathbf{y}]^\dagger] \\ &\quad - snr \mathbf{H}_1^\dagger \mathbb{E}_y [\mathbf{H}_2 \mathbf{P}_2 \mathbb{E}_{x_2|y} [\mathbf{x}_2 | \mathbf{y}] \mathbb{E}_{x_1|y} [\mathbf{x}_1 | \mathbf{y}]^\dagger] \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{\partial I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y})}{\partial \mathbf{P}_1^\dagger} &= snr \mathbf{H}_1^\dagger \mathbf{H}_1 \mathbf{P}_1 \mathbb{E}_{x_1} [\mathbf{x}_1 \mathbf{x}_1^\dagger] - snr \mathbf{H}_1^\dagger \mathbf{H}_1 \mathbf{P}_1 \mathbb{E}_y [\mathbb{E}_{x_1|y} [\mathbf{x}_1 | \mathbf{y}] \mathbb{E}_{x_1|y} [\mathbf{x}_1 | \mathbf{y}]^\dagger] \\ &\quad - snr \mathbf{H}_1^\dagger \mathbf{H}_2 \mathbf{P}_2 \mathbb{E}_y [\mathbb{E}_{x_2|y} [\mathbf{x}_2 | \mathbf{y}] \mathbb{E}_{x_1|y} [\mathbf{x}_1 | \mathbf{y}]^\dagger] \end{aligned} \quad (132)$$

$$\frac{\partial I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y})}{\partial \mathbf{P}_1^\dagger} = snr \mathbf{H}_1^\dagger \mathbf{H}_1 \mathbf{P}_1 \mathbf{E}_1 - snr \mathbf{H}_1^\dagger \mathbf{H}_2 \mathbf{P}_2 \mathbb{E}_y [\mathbb{E}_{x_2|y} [\mathbf{x}_2 | \mathbf{y}] \mathbb{E}_{x_1|y} [\mathbf{x}_1 | \mathbf{y}]^\dagger] \quad (133)$$

Similarly, we can derive the gradient of the mutual information in terms of the channel matrix that corresponds to the second user in the two-user MAC.

$$\frac{\partial I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y})}{\partial \mathbf{P}_2^\dagger} = snr \mathbf{H}_2^\dagger \mathbf{H}_2 \mathbf{P}_2 \mathbf{E}_2 - snr \mathbf{H}_2^\dagger \mathbf{H}_1 \mathbf{P}_1 \mathbb{E}_y [\mathbb{E}_{x_1|y} [\mathbf{x}_1 | \mathbf{y}] \mathbb{E}_{x_2|y} [\mathbf{x}_2 | \mathbf{y}]^\dagger] \quad (134)$$

Therefore, the gradient of the mutual information with respect to per user precoding and the per user MMSE and input estimates (or covariances) is as follows:

$$\nabla_{\mathbf{P}_1} I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}) = snr \mathbf{H}_1^\dagger \mathbf{H}_1 \mathbf{P}_1 \mathbf{E}_1 - snr \mathbf{H}_1^\dagger \mathbf{H}_2 \mathbf{P}_2 \mathbb{E}_y [\hat{\mathbf{x}}_2 \hat{\mathbf{x}}_1^\dagger] \quad (135)$$

$$\nabla_{\mathbf{P}_2} I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}) = snr \mathbf{H}_2^\dagger \mathbf{H}_2 \mathbf{P}_2 \mathbf{E}_2 - snr \mathbf{H}_2^\dagger \mathbf{H}_1 \mathbf{P}_1 \mathbb{E}_y [\hat{\mathbf{x}}_1 \hat{\mathbf{x}}_2^\dagger] \quad (136)$$

Therefore, Theorem 4 has been proved.

XII. APPENDIX D: PROOF OF THEOREM 5

From the steps in Theorem 3, we can see that:

$$\begin{aligned} p_y(\mathbf{y}) (\mathbf{y} - \sqrt{snr} \mathbf{H}_1 \mathbf{P}_1 \mathbb{E}_{x_1|y}[\mathbf{x}_1|y] - \sqrt{snr} \mathbf{H}_2 \mathbf{P}_2 \mathbb{E}_{x_2|y}[\mathbf{x}_2|y]) \\ = \mathbb{E}_{x_1, x_2} [p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) (\mathbf{y} - \sqrt{snr} \mathbf{H}_1 \mathbf{P}_1 \mathbf{x}_1 - \sqrt{snr} \mathbf{H}_2 \mathbf{P}_2 \mathbf{x}_2)] \end{aligned} \quad (137)$$

Therefore,

$$\begin{aligned} \sqrt{snr} \mathbf{H}_1 \mathbf{P}_1 \mathbb{E}_{x_1|y}[\mathbf{x}_1|y] - \sqrt{snr} \mathbf{H}_2 \mathbf{P}_2 \mathbb{E}_{x_2|y}[\mathbf{x}_2|y] \\ = \mathbf{y} - \frac{\mathbb{E}_{x_1, x_2} [p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) (\mathbf{y} - \sqrt{snr} \mathbf{H}_1 \mathbf{P}_1 \mathbf{x}_1 - \sqrt{snr} \mathbf{H}_2 \mathbf{P}_2 \mathbf{x}_2)]}{p_y(\mathbf{y})} \\ = \mathbf{y} + \frac{\mathbb{E}_{x_1, x_2} [\nabla_{\mathbf{y}} p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2)]}{p_y(\mathbf{y})} \\ = \mathbf{y} + \frac{\nabla_{\mathbf{y}} \mathbb{E}_{x_1, x_2} [p_{y|x_1, x_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2)]}{p_y(\mathbf{y})} \end{aligned} \quad (138)$$

Thus,

$$\sqrt{snr} \mathbf{H}_1 \mathbf{P}_1 \mathbb{E}[\mathbf{x}_1|y] - \sqrt{snr} \mathbf{H}_2 \mathbf{P}_2 \mathbb{E}[\mathbf{x}_2|y] = \mathbf{y} + \frac{\nabla_{\mathbf{y}} p_y(\mathbf{y})}{p_y(\mathbf{y})} \quad (139)$$

Therefore, Theorem 5 has been proved.

XIII. APPENDIX E: PROOF OF THEOREM 6

The MMSE Wiener filters are known to be MSE minimizers. In fact, this kind of receive filters inherently include the linear MMSE matrix. Therefore, to start the proof of the theorem, let's first define the linear MMSE as:

$$mmse(snr) = \underbrace{\mathbb{E}_{\mathbf{y}} [\|\mathbf{H}_1 \mathbf{P}_1 (\mathbf{x}_1 - \hat{\mathbf{x}}_1)\|^2]}_{\text{MMSE1}} + \underbrace{\mathbb{E}_{\mathbf{y}} [\|\mathbf{H}_2 \mathbf{P}_2 (\mathbf{x}_2 - \hat{\mathbf{x}}_2)\|^2]}_{\text{MMSE2}}, \quad (140)$$

with the linear estimates should be found by multiplying the received vector \mathbf{y} by the receive filter \mathbf{H}_r ,

$$\hat{\mathbf{x}}_1 = \mathbf{H}_{r1} \mathbf{y} \quad (141)$$

$$\hat{\mathbf{x}}_2 = \mathbf{H}_{r2} \mathbf{y} \quad (142)$$

We break down the problem into two parts for each per user MMSE. Notice that we call it MMSE since our target is the minimum MSE, therefore, we first take the first part, i.e., MSE1 and break it down; it follows that:

$$\tilde{\mathbf{E}}_1 = \mathbb{E}_{\mathbf{y}} [\|\mathbf{H}_1 \mathbf{P}_1 (\mathbf{x}_1 - \hat{\mathbf{x}}_1)\|^2] \quad (143)$$

$$= \mathbb{E}_{\mathbf{y}} [\mathbf{H}_1 \mathbf{P}_1 (\mathbf{x}_1 - \hat{\mathbf{x}}_1) (\mathbf{x}_1 - \hat{\mathbf{x}}_1)^\dagger (\mathbf{H}_1 \mathbf{P}_1)^\dagger] \quad (144)$$

$$= \mathbb{E}_{\mathbf{y}} [(\mathbf{H}_1 \mathbf{P}_1 \mathbf{x}_1 - \mathbf{H}_1 \mathbf{P}_1 \hat{\mathbf{x}}_1) (\mathbf{H}_1 \mathbf{P}_1 \mathbf{x}_1 - \mathbf{H}_1 \mathbf{P}_1 \hat{\mathbf{x}}_1)^\dagger] \quad (145)$$

$$\begin{aligned} &= \underbrace{\mathbb{E}_{\mathbf{x}_1} [\mathbf{H}_1 \mathbf{P}_1 \mathbf{x}_1 \mathbf{x}_1^\dagger (\mathbf{H}_1 \mathbf{P}_1)^\dagger]}_{\text{first term}} - \underbrace{\mathbb{E}_{\mathbf{x}_1} [\mathbf{H}_1 \mathbf{P}_1 \mathbf{x}_1 \hat{\mathbf{x}}_1^\dagger (\mathbf{H}_1 \mathbf{P}_1)^\dagger]}_{\text{second term}} \\ &\quad - \underbrace{\mathbb{E}_{\mathbf{x}_1} [\mathbf{H}_1 \mathbf{P}_1 \hat{\mathbf{x}}_1 \mathbf{x}_1^\dagger (\mathbf{H}_1 \mathbf{P}_1)^\dagger]}_{\text{third term}} + \underbrace{\mathbb{E}_{\mathbf{x}_1} [\mathbf{H}_1 \mathbf{P}_1 \hat{\mathbf{x}}_1 \hat{\mathbf{x}}_1^\dagger (\mathbf{H}_1 \mathbf{P}_1)^\dagger]}_{\text{forth term}} \end{aligned} \quad (146)$$

Lets work on each term in (146) separately. Notice that $\mathbb{E}_{\mathbf{x}_1} [\mathbf{x}_1 \mathbf{x}_1^\dagger] = \mathbb{E}_{\mathbf{x}_2} [\mathbf{x}_2 \mathbf{x}_2^\dagger] = \mathbf{I}$, $\mathbb{E}_{\mathbf{x}_1} [\mathbf{x}_1 \mathbf{x}_1^\mathbf{T}] = \mathbb{E}_{\mathbf{x}_2} [\mathbf{x}_2 \mathbf{x}_2^\mathbf{T}] = \mathbf{0}$, $\mathbb{E}_{x_1, x_2} [\mathbf{x}_2 \mathbf{x}_1^\dagger] = \mathbb{E}_{\mathbf{x}_1, \mathbf{x}_2} [\mathbf{x}_1 \mathbf{x}_2^\dagger] = \mathbf{0}$, $\mathbb{E}_n [\mathbf{n} \mathbf{n}^\dagger] = \mathbf{I}$, and $\mathbb{E}_n [\mathbf{n} \mathbf{n}^\mathbf{T}] = \mathbf{0}$.

The first term is:

$$\mathbb{E}_{\mathbf{x}_1} [(\mathbf{H}_1 \mathbf{P}_1 \mathbf{x}_1 \mathbf{x}_1^\dagger (\mathbf{H}_1 \mathbf{P}_1)^\dagger)] = \mathbb{E}_{\mathbf{x}_1} [\mathbf{H}_1 \mathbf{P}_1 (\mathbf{H}_1 \mathbf{P}_1)^\dagger], \quad (147)$$

The second term is:

$$\mathbb{E}_{\mathbf{x}_1} [\mathbf{H}_1 \mathbf{P}_1 \mathbf{x}_1 \hat{\mathbf{x}}_1^\dagger (\mathbf{H}_1 \mathbf{P}_1)^\dagger] = \mathbb{E}_{\mathbf{x}_1} [\mathbf{H}_1 \mathbf{P}_1 \mathbf{x}_1 (\mathbf{H}_{r1} \mathbf{y})^\dagger (\mathbf{H}_1 \mathbf{P}_1)^\dagger] \quad (148)$$

$$\mathbb{E}_{\mathbf{x}_1} [\mathbf{H}_1 \mathbf{P}_1 \mathbf{x}_1 \mathbf{y}^\dagger \mathbf{H}_{r1}^\dagger (\mathbf{H}_1 \mathbf{P}_1)^\dagger] = \mathbb{E}_{\mathbf{x}_1} [\mathbf{H}_1 \mathbf{P}_1 \mathbf{x}_1 (\mathbf{H}_1 \mathbf{P}_1 \mathbf{x}_1 + \mathbf{H}_2 \mathbf{P}_2 \mathbf{x}_2 + \mathbf{n})^\dagger \mathbf{H}_{r1}^\dagger (\mathbf{H}_1 \mathbf{P}_1)^\dagger] \quad (149)$$

$$= \mathbb{E}_{\mathbf{x}_1} [\mathbf{H}_1 \mathbf{P}_1 \mathbf{P}_1^\dagger \mathbf{H}_1^\dagger \mathbf{H}_{r1}^\dagger (\mathbf{H}_1 \mathbf{P}_1)^\dagger] \quad (150)$$

The third term is:

$$\mathbb{E}_{\mathbf{x}_1} [\mathbf{H}_1 \mathbf{P}_1 \hat{\mathbf{x}}_1 \mathbf{x}_1^\dagger (\mathbf{H}_1 \mathbf{P}_1)^\dagger] = \mathbb{E}_{\mathbf{x}_1} [\mathbf{H}_1 \mathbf{P}_1 \mathbf{H}_{r1} \mathbf{y} \mathbf{x}_1^\dagger (\mathbf{H}_1 \mathbf{P}_1)^\dagger] \quad (151)$$

$$= \mathbb{E}_{\mathbf{x}_1} [\mathbf{H}_1 \mathbf{P}_1 \mathbf{H}_{r1} \mathbf{H}_1 \mathbf{P}_1 (\mathbf{H}_1 \mathbf{P}_1)^\dagger] \quad (152)$$

The forth term is:

$$\mathbb{E}_{\mathbf{x}_1} \left[\mathbf{H}_1 \mathbf{P}_1 \widehat{\mathbf{x}}_1 \widehat{\mathbf{x}}_1^\dagger (\mathbf{H}_1 \mathbf{P}_1)^\dagger \right] = \mathbb{E}_{\mathbf{y}} \left[\mathbf{H}_1 \mathbf{P}_1 \mathbf{H}_{r1} \mathbf{y} \mathbf{y}^\dagger \mathbf{H}_{r1}^\dagger (\mathbf{H}_1 \mathbf{P}_1)^\dagger \right] \quad (153)$$

$$\begin{aligned} \mathbb{E}_{\mathbf{y}} \left[\mathbf{H}_1 \mathbf{P}_1 \mathbf{H}_{r1} \mathbf{y} \mathbf{y}^\dagger \mathbf{H}_{r1}^\dagger (\mathbf{H}_1 \mathbf{P}_1)^\dagger \right] &= \mathbb{E}_{\mathbf{y}} \left[\mathbf{H}_1 \mathbf{P}_1 \mathbf{H}_{r1} \mathbf{H}_1 \mathbf{P}_1 \mathbf{P}_1^\dagger \mathbf{H}_1^\dagger \mathbf{H}_{r1}^\dagger (\mathbf{H}_1 \mathbf{P}_1)^\dagger \right] \\ &+ \mathbb{E}_{\mathbf{y}} \left[\mathbf{H}_1 \mathbf{P}_1 \mathbf{H}_{r1} \mathbf{H}_2 \mathbf{P}_2 \mathbf{P}_2^\dagger \mathbf{H}_2^\dagger \mathbf{H}_{r1}^\dagger (\mathbf{H}_1 \mathbf{P}_1)^\dagger \right] + \mathbb{E}_{\mathbf{y}} \left[\mathbf{H}_1 \mathbf{P}_1 \mathbf{H}_{r1} \mathbf{H}_{r1}^\dagger (\mathbf{H}_1 \mathbf{P}_1)^\dagger \right] \end{aligned} \quad (154)$$

Therefore, the MSE1 can be given as follows:

$$\begin{aligned} MSE1(snr) &= Tr \left\{ \mathbf{H}_1 \mathbf{P}_1 (\mathbf{H}_1 \mathbf{P}_1)^\dagger \right\} - Tr \left\{ \mathbf{H}_1 \mathbf{P}_1 \mathbf{P}_1^\dagger \mathbf{H}_1^\dagger \mathbf{H}_{r1}^\dagger (\mathbf{H}_1 \mathbf{P}_1)^\dagger \right\} \\ &- Tr \left\{ \mathbf{H}_1 \mathbf{P}_1 \mathbf{H}_{r1} \mathbf{H}_1 \mathbf{P}_1 (\mathbf{H}_1 \mathbf{P}_1)^\dagger \right\} + Tr \left\{ \mathbf{H}_1 \mathbf{P}_1 \mathbf{H}_{r1} \mathbf{H}_1 \mathbf{P}_1 \mathbf{P}_1^\dagger \mathbf{H}_1^\dagger \mathbf{H}_{r1}^\dagger (\mathbf{H}_1 \mathbf{P}_1)^\dagger \right\} \\ &+ Tr \left\{ \mathbf{H}_1 \mathbf{P}_1 \mathbf{H}_{r1} \mathbf{H}_2 \mathbf{P}_2 \mathbf{P}_2^\dagger \mathbf{H}_2^\dagger \mathbf{H}_{r1}^\dagger (\mathbf{H}_1 \mathbf{P}_1)^\dagger \right\} + Tr \left\{ \mathbf{H}_1 \mathbf{P}_1 \mathbf{H}_{r1} \mathbf{H}_{r1}^\dagger (\mathbf{H}_1 \mathbf{P}_1)^\dagger \right\} \end{aligned} \quad (155)$$

We need now to find the minimum MSE1, and apply the KKT conditions to derive the optimal MMSE1

Wiener receive filter \mathbf{H}_{r1} as follows:

$$\frac{\partial MSE1}{\partial H_{r1}} = 0 \quad (156)$$

Capitalizing on the derivative rules for the trace of matrices, $\frac{\partial Tr(ABA^\dagger)}{\partial A} = AB^\dagger$, $\frac{\partial Tr(AB)}{\partial A} = B^\dagger$, and $\frac{\partial Tr(BA^\dagger)}{\partial A} = B$. We have:

$$\mathbf{H}_{r1}^* = \mathbf{P}_1^\dagger \mathbf{H}_1^\dagger (\mathbf{I} + \mathbf{P}_1^\dagger \mathbf{H}_1^\dagger \mathbf{P}_1 \mathbf{H}_1 + \mathbf{P}_2^\dagger \mathbf{H}_2^\dagger \mathbf{P}_2 \mathbf{H}_2)^{-1} \quad (157)$$

Similar steps to break down MSE2,

$$\tilde{\mathbf{E}}_2 = \mathbb{E}_{\mathbf{y}} \left[\|\mathbf{H}_2 \mathbf{P}_2 (\mathbf{x}_2 - \widehat{\mathbf{x}}_2)\|^2 \right], \quad (158)$$

leads to the optimal MMSE2 Wiener filter \mathbf{H}_{r2} as follows:

$$\mathbf{H}_{r2}^* = \mathbf{P}_2^\dagger \mathbf{H}_2^\dagger (\mathbf{I} + \mathbf{P}_1^\dagger \mathbf{H}_1^\dagger \mathbf{P}_1 \mathbf{H}_1 + \mathbf{P}_2^\dagger \mathbf{H}_2^\dagger \mathbf{P}_2 \mathbf{H}_2)^{-1} \quad (159)$$

Therefore, to estimate the user inputs linearly we do the following:

$$\widehat{\mathbf{x}}_1 = \mathbf{P}_1^\dagger \mathbf{H}_1^\dagger (\mathbf{I} + \mathbf{P}_1^\dagger \mathbf{H}_1^\dagger \mathbf{P}_1 \mathbf{H}_1 + \mathbf{P}_2^\dagger \mathbf{H}_2^\dagger \mathbf{P}_2 \mathbf{H}_2)^{-1} \mathbf{y} \quad (160)$$

$$\widehat{\mathbf{x}}_2 = \mathbf{P}_2^\dagger \mathbf{H}_2^\dagger (\mathbf{I} + \mathbf{P}_1^\dagger \mathbf{H}_1^\dagger \mathbf{P}_1 \mathbf{H}_1 + \mathbf{P}_2^\dagger \mathbf{H}_2^\dagger \mathbf{P}_2 \mathbf{H}_2)^{-1} \mathbf{y} \quad (161)$$

Therefore, Theorem 6 has been proved.

XIV. APPENDIX F: PROOF OF THEOREM 8

The possible solutions to (2) subject to (3) and (4) are characterized by the KKT conditions, which give necessary conditions for the matrix to be a critical point, known as the KKT or first-order conditions. To compute the KKT conditions, we first build the Lagrangian:

$$\mathcal{L}(\mathbf{P}_1, \mathbf{P}_2, \lambda_1, \lambda_2) = -I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}) - \lambda_1(Q_1 - \mathbf{P}_1) - \lambda_2(Q_2 - \mathbf{P}_2) - \mu_1\mathbf{P}_1 - \mu_2\mathbf{P}_2 \quad (162)$$

With primal feasibility conditions, $\lambda_1(Q_1 - \mathbf{P}_1) = 0$, $\mu_1\mathbf{P}_1 = 0$, $\lambda_2(Q_2 - \mathbf{P}_2) = 0$, and $\mu_2\mathbf{P}_2 = 0$, and dual feasibility conditions, $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$. In which the Lagrange multipliers λ_1 and λ_2 accounting for the in-equality constraint, has to be non-negative. The first-order conditions are given by:

$$\nabla_{\mathbf{P}_1} \mathcal{L}(\mathbf{P}_1, \mathbf{P}_2, \lambda_1, \lambda_2) = -\nabla_{\mathbf{P}_1} I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}) + \lambda_1 \mathbf{P}_1 = 0 \quad (163)$$

$$\nabla_{\mathbf{P}_2} \mathcal{L}(\mathbf{P}_1, \mathbf{P}_2, \lambda_1, \lambda_2) = -\nabla_{\mathbf{P}_2} I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}) + \lambda_2 \mathbf{P}_2 = 0 \quad (164)$$

Therefore, the solution of the optimal precoders satisfies:

$$\mathbf{P}_1^* = \nu_1^{-1} \mathbf{H}_1^\dagger \mathbf{H}_1 \mathbf{P}_1^* \mathbf{E}_1 - \nu_1^{-1} \mathbf{H}_1^\dagger \mathbf{H}_2 \mathbf{P}_2^* \mathbb{E}[\hat{\mathbf{x}}_2 \hat{\mathbf{x}}_1^\dagger] \quad (165)$$

$$\mathbf{P}_2^* = \nu_2^{-1} \mathbf{H}_2^\dagger \mathbf{H}_2 \mathbf{P}_2^* \mathbf{E}_2 - \nu_2^{-1} \mathbf{H}_2^\dagger \mathbf{H}_1 \mathbf{P}_1^* \mathbb{E}[\hat{\mathbf{x}}_1 \hat{\mathbf{x}}_2^\dagger] \quad (166)$$

where the solution follows from the gradient of the mutual information with respect to the precoding matrices. We eliminate λ_1 and λ_2 by introducing $\nu_1 = \frac{\lambda_1}{snr}$, and $\nu_2 = \frac{\lambda_2}{snr}$. Therefore, Theorem 8 has been proved.

XV. APPENDIX G: PROOF OF THEOREM 9

Digging into the depth of equations (39) and (40), we can do a singular value decomposition of the channel matrix $\mathbf{H}_1 = \mathbf{U}_{\mathbf{H}_1} \mathbf{\Lambda}_{\mathbf{H}_1} \mathbf{V}_{\mathbf{H}_1}^\dagger$, and $\mathbf{H}_2 = \mathbf{U}_{\mathbf{H}_2} \mathbf{\Lambda}_{\mathbf{H}_2} \mathbf{V}_{\mathbf{H}_2}^\dagger$, and the eigen value decomposition of the MMSE matrices, $\mathbf{E}_1 = \mathbf{U}_{\mathbf{E}_1} \mathbf{\Lambda}_{\mathbf{E}_1} \mathbf{V}_{\mathbf{E}_1}^\dagger$, and $\mathbf{E}_2 = \mathbf{U}_{\mathbf{E}_2} \mathbf{\Lambda}_{\mathbf{E}_2} \mathbf{V}_{\mathbf{E}_2}^\dagger$, we assume that the second terms in

equations (37) and (38) are zeros. Therefore, multiplying the first term of (37) by $\mathbf{P}_1^{\star\dagger}$, and of (38) by $\mathbf{P}_2^{\star\dagger}$, and taking the trace, which corresponds to each of the users power constraints. It follows that:

$$\begin{aligned}
Tr \{ \mathbf{P}_1^{\star} \mathbf{P}_1^{\star\dagger} \} &= Tr \{ \mathbf{H}_1 \mathbf{H}_1^\dagger \mathbf{P}_1^{\star} \mathbf{E}_1 \mathbf{P}_1^{\star\dagger} \} \\
&= Tr \{ \mathbf{P}_1^{\star\dagger} \mathbf{H}_1 \mathbf{H}_1^\dagger \mathbf{P}_1^{\star} \mathbf{E}_1 \} \\
&= Tr \{ \mathbf{E}_1 \mathbf{P}_1^{\star\dagger} \mathbf{H}_1 \mathbf{H}_1^\dagger \mathbf{P}_1^{\star} \} \\
&= Tr \{ \mathbf{P}_1^{\star} \mathbf{E}_1 \mathbf{P}_1^{\star\dagger} \mathbf{H}_1 \mathbf{H}_1^\dagger \} \\
&= Tr \{ \Lambda_{\mathbf{P}_1} \Lambda_{\mathbf{E}_1} \Lambda_{\mathbf{P}_1} \Lambda_{\mathbf{H}_1}^2 \} \\
&= Tr \{ \Lambda_{\mathbf{H}_1}^2 \Lambda_{\mathbf{P}_1} \Lambda_{\mathbf{E}_1} \Lambda_{\mathbf{P}_1} \} \\
&= Tr \{ \Lambda_{\mathbf{H}_1}^2 \Lambda_{\mathbf{P}_1}^2 \Pi \Lambda_{\mathbf{E}_1} \}
\end{aligned} \tag{167}$$

With:

$$\mathbf{U}_1 = \mathbf{V}_{\mathbf{H}_1}, \tag{168}$$

$$\mathbf{U}_2 = \mathbf{V}_{\mathbf{H}_2}, \tag{169}$$

$$\mathbf{D}_1 = \text{diag}(\sqrt{\mathbf{p}_{1,1}}, \dots, \sqrt{\mathbf{p}_{1,n_t}}), \tag{170}$$

$$\mathbf{D}_2 = \text{diag}(\sqrt{\mathbf{p}_{2,1}}, \dots, \sqrt{\mathbf{p}_{2,n_t}}), \tag{171}$$

$$\mathbf{R}_1 = \Pi \mathbf{U}_{\mathbf{E}_1}, \tag{172}$$

$$\mathbf{R}_2 = \Pi \mathbf{U}_{\mathbf{E}_2}. \tag{173}$$

Similar steps to prove the setup of the fixed point equation of the second user. Therefore, the fixed point equation of the per user optimal precoder precluding the other users interference is given as:

$$\mathbf{P}_1 = \mathbf{U}_1 \mathbf{D}_1 \mathbf{R}_1^\dagger \tag{174}$$

$$\mathbf{P}_2 = \mathbf{U}_2 \mathbf{D}_2 \mathbf{R}_2^\dagger \tag{175}$$

Therefore, Theorem 9 is proved. Note that we provide another proof for this theorem in [15].

XVI. APPENDIX H: PROOF OF THEOREM 10

The Lagrangian for the optimization problem (59) subject to (60) to (62) is given by:

$$\mathcal{L}(\mathbf{P}_1, \mathbf{P}_2, \lambda_1, \lambda_2) = -I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}) - \lambda_1(1 - \sum_{j=1}^{n_t} p_{1j}) - \lambda_2(1 - \sum_{j=1}^{n_t} p_{2j}) - \sum_{j=1}^{n_t} \mu_1 p_{1j} - \sum_{j=1}^{n_t} \mu_2 p_{2j} \quad (176)$$

With primal feasibility conditions, $\lambda_1(1 - \sum_{j=1}^{n_t} p_{1j}) = 0$, $\mu_1 \sum_{j=1}^{n_t} p_{1j} = 0$, $\lambda_2(1 - \sum_{j=1}^{n_t} p_{2j}) = 0$, and $\mu_2 \sum_{j=1}^{n_t} p_{2j} = 0$, and dual feasibility conditions, $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$. The KKT conditions are given by:

$$\frac{\partial \mathcal{L}(\mathbf{P}_1, \mathbf{P}_2, \lambda_1, \lambda_2)}{\partial \mathbf{p}_1} = -\frac{\partial I(x_1, x_2; y)}{\partial \mathbf{p}_1} + \lambda_1 - \mu_1 = 0 \quad (177)$$

$$\frac{\partial \mathcal{L}(\mathbf{P}_1, \mathbf{P}_2, \lambda_1, \lambda_2)}{\partial \mathbf{p}_2} = -\frac{\partial I(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y})}{\partial \mathbf{p}_2} + \lambda_2 - \mu_2 = 0 \quad (178)$$

Note that $\mu_1 = 0$ for $\mathbf{p}_1 \succ \mathbf{0}$ and $\mu_2 = 0$ for $\mathbf{p}_2 \succ \mathbf{0}$. Therefore, the derivative with respect to \mathbf{p}_1 , and \mathbf{p}_2 respectively, is given by:

$$\frac{\partial \mathcal{L}(\mathbf{P}_1, \mathbf{P}_2, \lambda_1, \lambda_2)}{\partial \mathbf{p}_1} = \frac{snr}{\sqrt{\mathbf{p}_1}} \left(\mathbf{P}_1^* \mathbf{H}_1^\dagger \mathbf{H}_1 \mathbf{P}_1^* \mathbf{E}_1 - \mathbf{P}_1^* \mathbf{H}_1^\dagger \mathbf{H}_2 \mathbf{P}_2^* \mathbb{E}[\widehat{\mathbf{x}}_2 \widehat{\mathbf{x}}_1^\dagger] \right)_{1j} \quad (179)$$

$$\frac{\partial \mathcal{L}(\mathbf{P}_1, \mathbf{P}_2, \lambda_1, \lambda_2)}{\partial \mathbf{p}_2} = \frac{snr}{\sqrt{\mathbf{p}_2}} \left(\mathbf{P}_2^* \mathbf{H}_2^\dagger \mathbf{H}_2 \mathbf{P}_2^* \mathbf{E}_2 - \mathbf{P}_2^* \mathbf{H}_2^\dagger \mathbf{H}_1 \mathbf{P}_1^* \mathbb{E}[\widehat{\mathbf{x}}_1 \widehat{\mathbf{x}}_2^\dagger] \right)_{2j} \quad (180)$$

Therefore, the optimal power allocation satisfies:

$$\mathbf{p}_1^* = \gamma_1^{-1} \left(\mathbf{P}_1^* \mathbf{H}_1^\dagger \mathbf{H}_1 \mathbf{P}_1^* \mathbf{E}_1 - \mathbf{P}_1^* \mathbf{H}_1^\dagger \mathbf{H}_2 \mathbf{P}_2^* \mathbb{E}[\widehat{\mathbf{x}}_2 \widehat{\mathbf{x}}_1^\dagger] \right)_{1j} \quad (181)$$

$$\mathbf{p}_2^* = \gamma_2^{-1} \left(\mathbf{P}_2^* \mathbf{H}_2^\dagger \mathbf{H}_2 \mathbf{P}_2^* \mathbf{E}_2 - \mathbf{P}_2^* \mathbf{H}_2^\dagger \mathbf{H}_1 \mathbf{P}_1^* \mathbb{E}[\widehat{\mathbf{x}}_1 \widehat{\mathbf{x}}_2^\dagger] \right)_{2j} \quad (182)$$

where $\gamma_1 = \frac{\lambda_1}{snr}$, and $\gamma_2 = \frac{\lambda_2}{snr}$. Therefore, Theorem 10 has been proved.

REFERENCES

- [1] J. Cioffi, S. Jagannathan, M. Mohseni, and G. Ginis, "Cupon: the copper alternative to pon 100 gb/s dsl networks [accepted from open call]," *IEEE Communications Magazine*, vol. 45, no. 6, pp. 132–139, 2007.
- [2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. 2nd ed. New York: Wiley, 2006.
- [3] F. Perez-Cruz, M. R. D. Rodrigues, and S. Verdú, "MIMO Gaussian channels with arbitrary inputs: optimal precoding and power allocation," *IEEE Transactions on Information Theory*, vol. 56, no. 3, pp. 1070–1084, March 2010.
- [4] G. Kramer, "Interference channels handout," *IEEE Winter school of Information Theory, UPF and CTTC, Barcelona*, 2011.

- [5] J. Andrews, "Interference cancellation for cellular systems: a contemporary overview," *IEEE Transactions on Wireless Communications*, vol. 12, no. 2, pp. 19–29, 2005.
- [6] B. Rimoldi and R. Urbanke, "A rate-splitting approach to the Gaussian multiple-access channel," *IEEE Transactions on Information Theory*, vol. 42, no. 2, pp. 364–375, 1996.
- [7] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.
- [8] S. P. Weber, J. Andrews, X. Yang, and G. de Veciana, "Transmission capacity of wireless ad hoc networks with successive interference cancellation," *IEEE Transactions on Information Theory*, vol. 53, no. 8, pp. 2799–2814, 2007.
- [9] J. Blomer and N. Jindal, "Transmission capacity of wireless ad hoc networks: Successive interference cancellation vs. joint detection," in *IEEE International Conference on Communications (ICC)*, 2009, pp. 1–5.
- [10] D. P. Palomar and S. Verdu, "Gradient of mutual information in linear vector Gaussian channels," *IEEE Transactions on Information Theory*, vol. 52, No.1, pp.141-154, January 2006.
- [11] M. Lamarca, "Linear precoding for mutual information maximization in mimo systems," in *International Symposium on Wireless Communication Systems (ISWCS)*, 2009.
- [12] D. Guo and S. Shamai and S. Verdu, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Transactions on Information Theory*, vol. 51, pp. 1261–1282, April 2005.
- [13] A. Lozano, A. Tulino, and S. Verdu, "Optimum power allocation for parallel Gaussian channels with arbitrary input distributions," *IEEE Transactions on Information Theory*, vol. 52, no. 7, pp. 3033–3051, 2006.
- [14] S. A. M. Ghanem, "MAC Gaussian channels with arbitrary inputs: Optimal precoding and power allocation," in *IEEE International Conference on Wireless Communications and Signal Processing (WCSP)*, Huangshan, China, October 2012.
- [15] —, "Optimal power allocation and optimal precoding with multi-cell processing," *IEEE 77th Vehicular Technology Conference: VTC-Spring, Dresden, Germany*, June 2013.
- [16] —, *Analysis, Modeling, Design, and Optimization of Future Communications Systems: From Theory to Practice*, Ph.D. Thesis, December 2013.
- [17] —, "Mutual information for generalized arbitrary binary input constellations," in *MAP-Tele Workshop*, February 2010.
- [18] —, "Multi-cell processing with limited cooperation: A novel framework to timely designs and reduced feedback with general inputs," <http://arxiv.org/abs/1411.0594>, 2014.
- [19] R. Bustin, H. V. Poor, and S. Shamai, "The effect of maximal rate codes on the interfering message rate," *CoRR*, vol. abs/1404.6690, 2014.
- [20] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge University Press, 2004.
- [21] S. Verdu, "Spectral efficiency in the wideband regime," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1319–1343, 2002.